Feature Extraction Techniques for Grasp Classification

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Engineering at the University of Canterbury

---

Rodney B. Elliott

August 1998

Abstract

This thesis examines the ability of four signal parameterisation techniques to provide discriminatory information between six different classes of signal. This was done with a view to assessing the suitability of the four techniques for inclusion in the real-time control scheme of a next generation robotic prosthesis. Each class of signal correlates to a particular type of grasp that the robotic prosthesis is able to form.

Discrimination between the six classes of signal was done on the basis of parameters extracted from four channels of electromyographic (EMG) data that was recorded from muscles in the forearm. Human skeletal muscle tissue produces EMG signals whenever it contracts. Therefore, providing that the EMG signals of the muscles controlling the movements of the hand vary sufficiently when forming the different grasp types, discrimination between the grasps is possible. While it is envisioned that the chosen command discrimination system will be used by mid-forearm amputees to control a robotic prosthesis, the viability of the different parameterisation techniques was tested on data gathered from able-bodied volunteers in order to establish an upper limit of performance.

The muscles from which signals were recorded are: the extensor pollicis brevis and extensor pollicis longus pair (responsible for moving the thumb); the extensor communis digitorum (responsible for moving the middle and index fingers); and the extensor carpi ulnaris (responsible for moving the little finger).

The four signal parameterisation techniques that were evaluated are:

1. **Envelope Maxima.** This method parameterises each EMG signal by the maximum value of a smoothed fitted signal envelope. A tenth order polynomial is fitted to the rectified EMG signal peaks, and the maximum value of the polynomial is used to parameterise the signal.

2. **Orthogonal Decomposition.** This method uses a set of orthogonal functions to decompose the EMG signal into a finite set of orthogonal components. Each burst is then parameterised by the coefficients

of the set of orthogonal functions. Two sets of orthogonal functions were tested: the Legendre polynomials, and the wavelet packets associated with the scaling functions of the Haar wavelet (referred to as the Haar wavelet for brevity).

3. **Global Dynamical Model.** This method uses a discretised set of non-linear ordinary differential equations to model the dynamical processes that produced the recorded EMG signals. The coefficients of this model are then used to parameterise the EMG signal.

4. **EMG Histogram.** This method formulates a histogram detailing the frequency with which the EMG signal enters particular voltage bins, and uses these frequency measurements to parameterise the signal.

Ten sets of EMG data were gathered and processed to extract the desired parameters. Each data set consisted of 600 grasps—100 grasp records of four channels of EMG data for each of the six grasp classes. From this data a hit rate statistic was formed for each feature extraction technique. The mean hit rates obtained from the four signal parameterisation techniques that were tested are summarised in Table 1. The EMG histogram provided

Parameterisation Technique	Hit Rate (%)
Envelope Maxima	75
Legendre Polynomials	77
Haar Wavelets	79
Global Dynamical Model	75
EMG Histogram	81

Table 1: Hit Rate Summary.

the best mean hit rate of all the signal parameterisation techniques of 81%. However, like all of the signal parameterisations that were tested, there was considerable variance in hit rates between the ten sets of data. This has been attributed to the manner in which the electrodes used to record the EMG signals were positioned. By locating the muscles of interest more accurately, consistent hit rates of 95% are well within reach.

The fact that the EMG histogram produces the best mean hit rates is surprising given its relative simplicity. However, this simplicity makes the EMG histogram feature ideal for inclusion in a real-time control scheme.

Acknowledgements

There are a number of people to whom I am indebted for their help during the course of this thesis.

I would firstly like to thank to Dr Reg Dunlop for starting me on an interesting course of research.

I would like to extend my most sincere thanks to Dr Chris Damaren, for acting as my supervisor while Dr Dunlop was on sabbatical. Dr Damaren kindly took me under his wing, and through judicious waving of the carrot and wielding of the stick helped turn me into a more competent and disciplined researcher.

I would like to thank Andrew Cree for his help in all matters relating to computer hardware and software. Over the course of this research I asked Andy a vast number of questions. Some of them were stupid, and most of them were repeated, and yet Andy never lost his patience. Well, almost never!

I would like to warmly thank Scott Stringer of the Burwood Artificial Limb Centre for his help in all matters relating to prosthetic devices. Scott very kindly gave up his time to give me a guided tour of the Burwood Artificial Limb Centre, and also to answer my questions on numerous occasions.

I would also like to thank the volunteers who offered their time so that I could gather the data that I required. Knowing just how excruciatingly boring the procedure is from personal experience, I fully appreciate your patience.

Finally, I would like to thank the Brian Mason Scientific and Technical Trust for their generous financial support of this research. Without their help, this work would of progressed a great deal slower.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

History is littered with accounts of battles and natural disasters that resulted in the loss of life and limb. Until quite recently, incidents resulting in the latter usually also led to the former. Naval surgeons of the last century for instance used to boast at their ability to completely remove an entire arm or leg in less than fifteen seconds. This fifteen seconds no doubt felt like an eternity to the unfortunate sailor on whom this "operation" was performed—usually without the aid of anesthetic. After the limb had been hacked off, the resulting stump was plunged into a bucket of hot tar to seal the wound. Providing that the shock did not kill the patient outright, then there was a chance that they would live to fight another day.

Advances in the field of medicine in the last 50 or-so years have ensured that anyone facing a limb amputation today stands a greatly improved chance of surviving the procedure, barring any unforeseen complications. Advances in the field of prosthetics have also meant that amputees are able to resume a far more productive life than was once possible.

1.1 Current Prosthetic Technology

Every year between 10 and 15 New Zealanders [Stringer 1998] undergo a mid-forearm amputation[1] as a result of either congenital defect, accident, or disease. The healing time from such drastic surgery ranges from a remarkable three to five weeks for someone who is young and fit, up to a maximum of approximately eight weeks for those of lesser health or advancing age. After this initial recuperation period, an appointment is made on the patient's behalf by the surgeon at one of the five clinics run by the New Zealand

[1] See Appendix A for an interesting and detailed account of the surgical procedure used when performing such a mid-forearm amputation.

Artificial Limb Board. These five clinics are located in Auckland, Hamilton, Wellington, Christchurch and Dunedin and it is their responsibility to supply, fit and maintain a variety of artificial limbs. Mid-forearm amputees in New Zealand are currently fitted with one of two prosthetic devices: the prosthetic hook, or the myoelectric hand.

1.1.1 Prosthetic Hook

The prosthetic hook is a simple one degree of freedom gripper consisting of two curved aluminum alloy or stainless steel jaws. Into these jaws are bonded a set of neoprene grip inserts whose purpose is to aid with object acquisition and retention. The two jaws of the prosthetic hook are normally held closed by several powerful rubber bands (Figure 1.1). The prosthetic hook

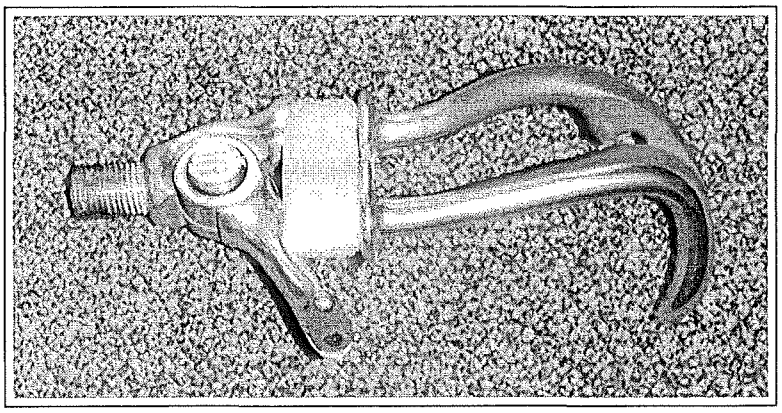

Figure 1.1: A Typical Prosthetic Hook.

is attached to the amputee via a fibreglass socket into which the amputee's stump is inserted. The prosthetic hook screws into a fitting in the end of the socket, and is prevented from unscrewing by a small grub screw. The socket is in turn held in place on the arm by either a triceps pad or epicondyle clip.

The opening and closing of the prosthetic hook is controlled by the amputee through the use of a shoulder harness. This is a synthetic webbing harness that forms a figure of eight when worn. As well as its function as a actuator, the shoulder harness is also ultimately responsible for securing the prosthetic hook and socket to the amputee via the triceps pad or epicondyle clip. Examination of Figure 1.2 reveals how the different components discussed thus far are assembled into a single functional unit. Note that Figure 1.2 shows a triceps pad being used rather than an epicondyle clip. If the workload that the amputee is predicted to subject the prosthetic hook

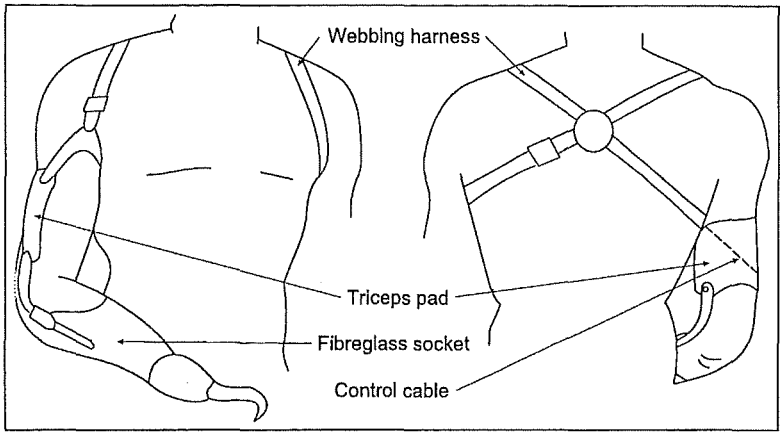

Figure 1.2: Prosthetic Hook Assembly.

to is light, then an epicondyle clip may be fitted instead. It consists of a metal clip that fits around the posterior of the epicondyle of the humerus, thus preventing the prosthetic hook and socket from falling off.

The opening and closing of the prosthetic hook is achieved through the rounding of the adjacent shoulder. By moving the adjacent shoulder forward, tension is applied to the stainless steel control cable attached to the shoulder harness (see Figure 1.2). The control cable is housed in a flexible metal sleeve and travels down the upper arm and along the socket until terminating at the thumb of the prosthetic hook (see Figure 1.2). This force applied to the thumb of the prosthetic hook counteracts that of the rubber bands and the jaws of the prosthetic hook open. Likewise, by moving the shoulder back, the rubber bands take up the slack in the control cable and the jaws close.

The great majority of mid forearm amputees in New Zealand are fitted with a prosthetic hook. They are extremely robust and immune to all but the harshest of treatments. They are also available in a variety of jaw configurations to suit most types of work. Their simple construction also makes them easy to repair and maintain, and at an average cost of $650 for the hook plus $2000 for the socket, harness, and fitting [Flanagan 1998], they are significantly cheaper than the myoelectric hand.

1.1.2 Myoelectric Hand

The alternative to the prosthetic hook is the myoelectric hand. The myoelectric hand is also a simple one degree of freedom gripper. It consists of two fingers and a thumb in opposition. These two sets of digits can be moved toward or away from each other in a manner similar to the jaws of

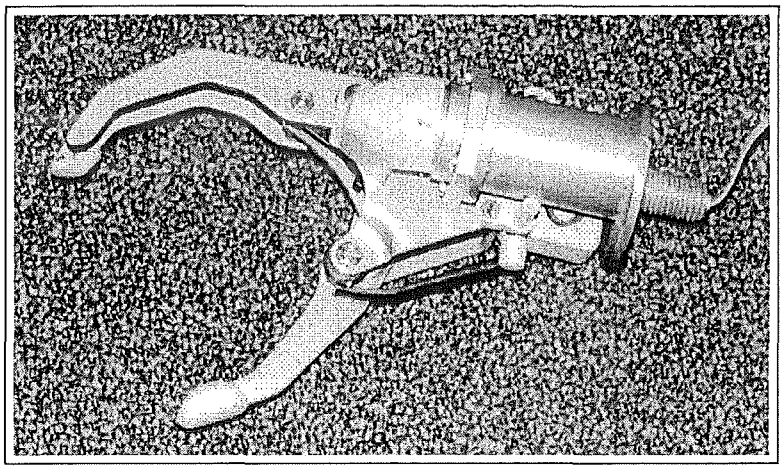

Figure 1.3: The Myoelectric Hand.

the prosthetic hook. Unlike the prosthetic hook however, this movement is controlled through an electro-mechanical system rather than manually (Figure 1.3). For a simple myoelectric hand, two sets of pressure sensitive electrodes are moulded into the socket to which the myoelectric hand attaches. By pushing the stump against one of these sets of electrodes, an electrical circuit is closed which opens the hand through a motor and reduction gearbox attached to the fingers and thumb of the myoelectric hand. Likewise, by pushing the stump against the other set of electrodes, the hand is able to be closed. Unlike the prosthetic hook which screws into the socket, the myoelectric hand is held in place either by a simple collar that can be tightened to set the myoelectric hand in a particular orientation, or by a quick release spline collar that enables the orientation to be changed whenever the amputee wishes. Along with the two sets of electrodes to control the opening and closing of the three digits, the socket of a myoelectric hand contains the battery packs, and may also contain a third set of electrodes that are used to control an optional powered wrist. In order to protect the workings of the myoelectric hand from damage, the entire mechanism is housed in a tough, anthropomorphically-correct rubber glove. This glove is shown in Figure 1.4. The shoulder harness used to keep the myoelectric hand and its accompanying socket attached to the arm of the amputee is identical to that used with the prosthetic hook, with the exception that the control cable is removed.

The much greater complexity of the myoelectric hand is reflected in its cost. The hand itself can cost up to $6000, in addition to which must be purchased two sets of electrodes at $1000 each, two battery packs at $600 each, and a battery charger ($1000). Add to this fitting costs of $1000 and it

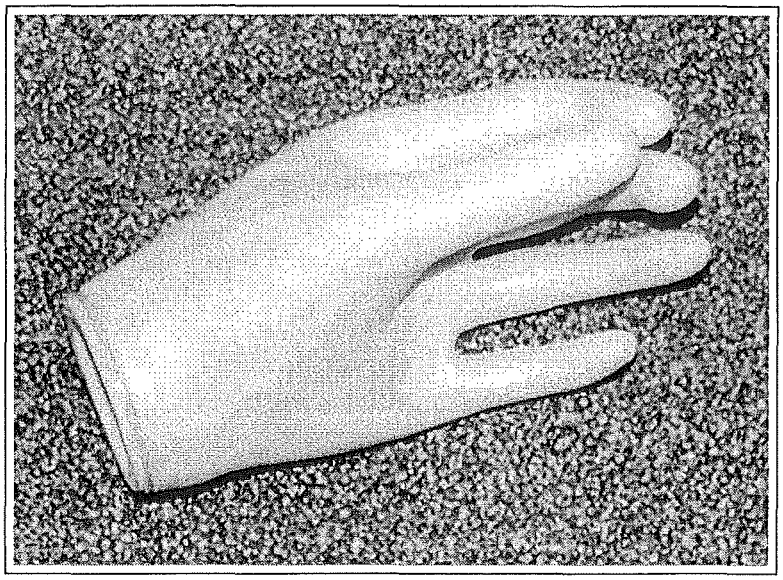

Figure 1.4: Protective Glove of the Myoelectric Hand.

becomes apparent that a myoelectric hand is significantly more costly than a prosthetic hook [Flanagan 1998]. This fact ensures that very few amputees are fitted with a myoelectric hand. The criteria for determining who will be fitted with a myoelectric hand rather than a prosthetic hook are quite simple [Stringer 1996]. All the prospective amputee has to do is to match one or more of the following criteria:

1. The amputee must be a child whose parents insist that they be fitted with a myoelectric hand for fear of the taunting that the child receive at school.

2. The amputee must be willing to constantly harass the New Zealand Artificial Limb Board until they relent and provide a myoelectric hand.

3. The amputee must be willing to make a significant contribution towards the cost of the myoelectric hand.

Overall, the myoelectric hand does not compare favourably with the prosthetic hook. It is far heavier and more expensive, and does not have anywhere near the load capacity of the prosthetic hook. In addition, the myoelectric hand lacks the force feedback that the manual control of the prosthetic hook provides. But despite these disadvantages, the myoelectric hand, along with the prosthetic hook, are being used to great effect by the amputee community. From digging the garden to preparing a meal, amputees are performing the

same tasks with the same degree of competence as their able-bodied counterparts. It must be stated however that this success is due more to human adaptability and perseverance than it is to any design features inherent in either the myoelectric hand or prosthetic hook.

1.2 Robotic Manipulator Review

In order to locate a more functional hand replacement it is necessary to look to the latest generation of robotic manipulators currently under development in research institutions and universities throughout the world.

For many years engineers and researchers have been attempting to design robotic manipulators that incorporate some of the abilities of the human hand. The human hand is an extraordinary device, capable of exerting forces varying by several orders of magnitude. It can pick up the smallest and most delicate of items without damage, as well as lifting objects weighing in excess of 1000 N. Its multiple digits, variable compliance, and redundant degrees of freedom enable the human hand to pick up, grasp, and manipulate a vast range of tools and objects. Its sensory system remains unmatched in its ability to detect force, slip, texture, temperature, and humidity. All of these features make the human hand the yardstick against which aspiring robotic manipulator designs must be measured.

The great advantage that the human hand has over robotic manipulators is its significantly longer evolution time. It is over the last several million years that the human hand has evolved to become the remarkable general purpose manipulator that it is today. Robotic manipulators have instead in their brief history tended to be far more specialised in terms of their function. For instance, industrial robots are designed to move large payloads, and have no use of the ability to perform fine movements. Conversely, electronic assembly robots are very adept at fine pick and place tasks, but are unable to lift even the smallest proportion of their own weight.

There are however tasks for which a robotic manipulator with the abilities of the human hand would be ideally suited. Tasks requiring remote teleoperation, where a robot is controlled by a human operator are an excellent example. Remote teleoperation is used in environments where it is either unwise or impractical for a person to be sent, and where the task is too unstructured for an autonomous robot. The handling of nuclear materials and the manipulation of objects in space are examples of situations where remote teleoperation is a viable control method. Another area in which a human hand-like robotic manipulator is very useful is in research involving the human hand itself. For instance, researching how the human hand is able

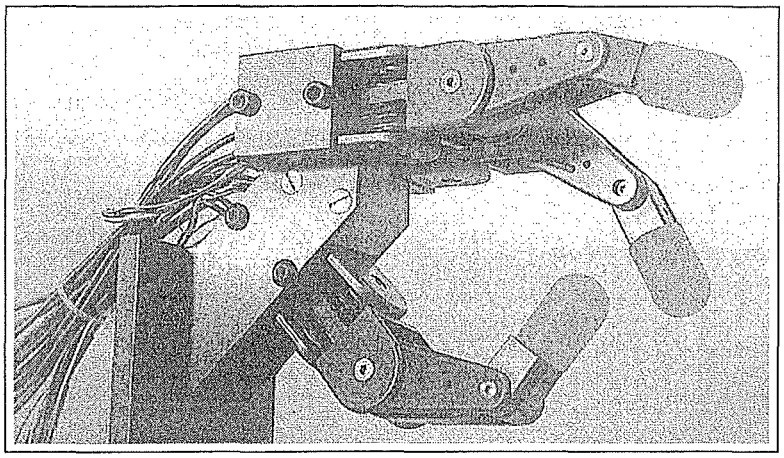

Figure 1.5: The Salisbury Hand. [Rosheim 1994, p. 200].

to dynamically manipulate an object without dropping or damaging it is best solved by a hand-like robotic manipulator on which different hypothesis may be tested.

But perhaps the most promising use for a human hand-like robotic manipulator is in the field of prosthetics. A robotic prosthetic hand would be a considerable improvement over the prosthetic hook and myoelectric hand in terms of its inherent functionality. It is the control of such a so-called next generation prosthesis which was the subject of the research upon which this thesis is based. Specifically, this thesis deals with the problem of developing a system which allows the amputee to command the movements of a next generation prosthetic hand. Before detailing this research, it is instructive to examine those anthropomorphic robotic manipulators available today which could be used as the starting point for the design of the next generation of prosthetic hands.

1.2.1 Salisbury Hand

The Salisbury Hand (Figure 1.5) consists of three, three degree of freedom digits, each controlled by four teflon cables. Although the three digits of the Salisbury Hand are identical, they have been physically configured into two fingers and an opposing thumb. The control cables are ducted through teflon coated conduits to an external actuator pack consisting of 12 DC servo motors, each with a 25:1 reduction gearbox [Rosheim 1994]. The Salisbury Hand was designed by Kenneth J. Salisbury as part of his doctoral research at Stanford University, and it is primarily used for research into fingertip

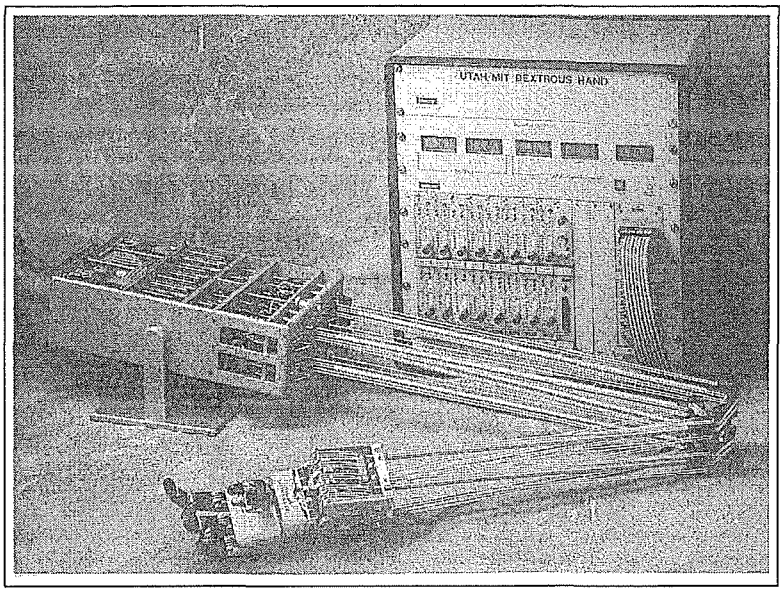

Figure 1.6: The Utah/MIT Hand, Drive, and Controller. [Rosheim 1994, p. 207].

prehension.

1.2.2 Utah/MIT Hand

The Utah/MIT Hand (Figure 1.6) was designed in the mid 1980's in order to study the problems of grasping and finger manipulation. It consists of three fingers and a thumb in opposition, with each digit having four degrees of freedom. As with the Salisbury Hand, the Utah/MIT Hand is actuated through a set of cables. These cables are controlled by a set of double acting glass pneumatic cylinders which, due to the design of the hand, have to be routed through an articulated arm-like frame [Rosheim 1994].

1.2.3 Belgrade/USC Hand

The Belgrade/USC Hand (Figure 1.7) was developed by the Electrical Engineering and Computer Science Departments at the University of Southern California, Los Angeles, and the Electrical Engineering Department at the University of Belgrade, Belgrade, Yugoslavia. The Belgrade/USC Hand is an anthropomorphically correct design with four fingers and an opposable thumb. The four fingers are driven by two motors, with each motor driving two adjacent fingers through a lead screw and differential whiffle tree.

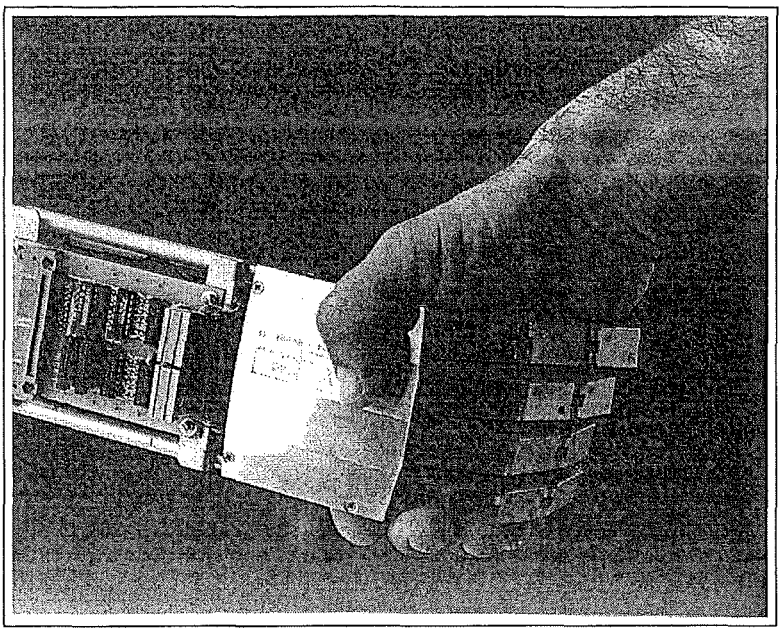

Figure 1.7: The Belgrade/USC Hand. [Rosheim 1994, p. 210].

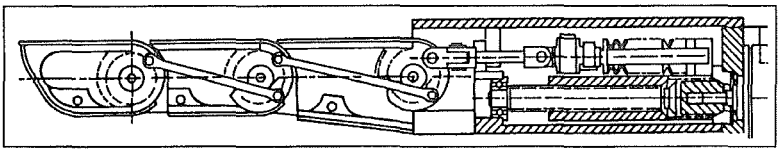

Figure 1.8: Sectional Side View of One of the Fingers of the Belgrade/USC Hand. [Rosheim 1994, p. 211].

Each finger has three joints, but the system of linkages that comprise the fingers results in these three joints only contributing one degree of freedom (Figure 1.8) to each finger. The whiffle tree itself does however introduce an additional degree of freedom, which is "shared" between adjacent fingers. This degree of freedom only comes into play when one of the fingers in an adjacent pair is in contact with an object, and results in the other finger being driven towards the object at twice normal closing speed. The thumb is also driven by two motors; one provides the mechanism by which the thumb is able to move into opposition, and the other closes the thumb through a mechanism similar to that used in the four fingers. The Belgrade/USC Hand was built to aid research in incorporating intelligence into hand shape control software [Rosheim 1994].

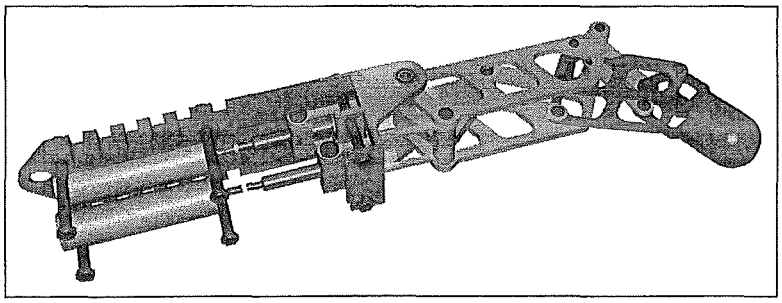

Figure 1.9: A Cutaway View of the UOC Finger.

1.2.4 The UOC Hand

The UOC (University of Canterbury) Hand is an ongoing project at the University of Canterbury, Christchurch, New Zealand. It was initiated as a result of discussions between Dr. G. R. Dunlop of the University of Canterbury, and Professor M. Vuskovic of San Diego State University on the operation and deficiencies of the Belgrade/USC Hand. A feasibility study was undertaken at the University of Canterbury to attempt to find the best strategy for improving the Belgrade/USC Hand. This study was carried out by Laurent Magnier and Hugues Monier [1993], students visiting from the Mechanical Engineering Department of the Ecole Nationale D'Ingenieurs, Saint-Etern, France. The feasibility study examined several improvements to the Belgrade/USC Hand, notably the repositioning of the finger joints to improve finger tip force, and the increase of each finger's degrees of freedom to two. The resulting improved Belgrade/USC finger was further modified [Traub 1996, Ward 1996], and a prototype built (Figure 1.9). It is this UOC Finger upon which the proposed UOC Hand is based [Dunlop and Ward 1995]. The proposed UOC Hand consists of six UOC Fingers, two of which act as opposable thumbs. In addition to this is a motor for spreading the fingers (Figure 1.10). Further work has since been carried out on the UOC Finger design. This consisted of optimising the joint positions of the UOC Finger using genetic algorithms [Bain 1998]. The purpose of this additional work was to provide the largest finger working area and finger tip force possible.

1.3 Command Discrimination

The robotic manipulators detailed in Section 1.2 above vary widely in their physical configuration. The number of digits, the degrees of freedom per digit, the method of actuation–all differ from one robotic manipulator to the

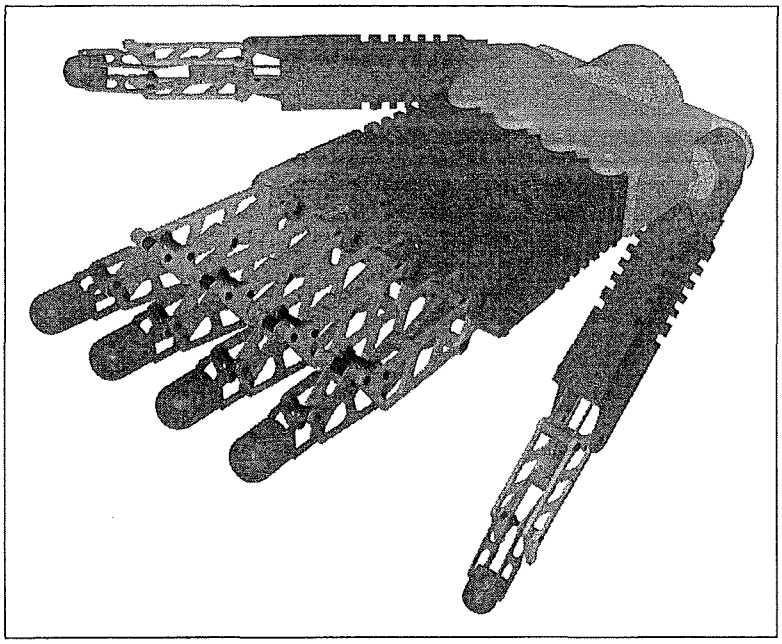

Figure 1.10: The Proposed UOC Hand.

next. When it comes to issues of control however, the problems faced are the same regardless of which robotic manipulator is chosen. The amount of processing power required, the robustness of the control algorithms, their suitability for real-time implementation—these are some of the factors that need to be taken into account when designing a control system for such a robotic manipulator. Such questions need to be answered regardless of whether the manipulator is to be used for research in the laboratory or remote teleoperation out in the field. The use of a robotic manipulator as a next generation prosthetic hand raises additional questions. Of paramount importance are those regarding the control system interface. Specifically, the acceptance of such a next generation prosthetic hand hinges on the user interface being as simple to use as possible. This is due to the fact that amputee's will be expecting to be able to use the robotic manipulator all day, every day of the year. If it is not easy to use then it will be rejected, no matter what improvements in functionality it promises.

The control of a robotic manipulator for use as a prosthetic hand can be split up into two quite distinct sections:

1. **Command Interpretation.** This part of the control scheme is responsible for determining what action it is that the user wishes the robotic manipulator to perform.

2. **Actioning of the Command.** This part of the control scheme is responsible for the actioning of the command given by the user. This may include aspects such as moving the manipulator into a particular configuration, control of grasp force, and the elimination of slip.

Control texts and journals are full of techniques with which the second point may be accomplished. The research described in this thesis sought to find a solution to the first point. Specifically, the research objective was to develop a seamless user interface which would allow selection of one of a number of commands for a robotic prosthesis with the greatest degree of accuracy possible. Before examining the techniques by which this was accomplished in detail, it is appropriate to examine the methods by which others in the field have sought to solve this problem.

1.4 Command Discriminator Review

A number of different techniques have been developed to enable a human operator to command the movements of a robotic prosthesis or manipulator. The great majority of these schemes use the human electromyographic (EMG) signal as their input. The EMG signal is a record of the electrical activity within a particular muscle, and its origin and function is covered in depth in Chapter 2. In order to make the problem of interpreting the users intentions more tractable, the number of possible commands that may be issued is limited. This is done both to aid the design engineer as well as the user themselves. It modifies the problem of interpreting the user's intentions into one of discriminating between a small number of possible commands. Of the different techniques used to perform this discrimination, the majority fall into one of two categories:

1. **Time Domain-Based Command Discrimination.** These techniques manipulate the EMG signals in the time domain in order to discriminate between signals from different hand commands.

2. **Frequency Domain-Based Command Discrimination.** These techniques manipulate the EMG signals in the frequency domain in order to discriminate between signals from different hand commands.

1.4.1 Time Domain-Based Command Discrimination

Vuskovic et al. [1995, 1996] developed a system enabling discrimination between six different grasp types: small cylinder, large cylinder, small sphere,

large sphere, small disk, and key. Four channels of EMG data were recorded from the muscles in the forearm responsible for controlling the movement of the thumb (extensor pollicis group), the index and middle fingers (extensor communis digitorum), and the little finger (extensor carpi ulnaris). This data was recorded as volunteers grasped each one of the six objects a large number of times. From each EMG data record a single feature was extracted from each channel. The feature chosen was the maxima of the processed EMG envelope, resulting in each data record being represented by a 4×1 feature vector. The extracted features from a number of grasps of a particular item will form a four dimensional hyperellipsoid or hypercluster. Maximum separation of the six hyperclusters was achieved by transforming the original feature space through the use of the principle component discriminant transform. The resulting maximal hypercluster mean separation results in the greatest possibility of a successful classification of an observation whose grasp identity is unknown. The Mahalanobis distance is used as the discriminator whereby an unknown observation is assigned the grasp identity of the hypercluster whose mean is closest, as measured using the Mahalanobis distance. Vuskovic also advocates the use of the principle component discriminant transform eigenvalues to rank the axes of the transformed feature space in terms of their discriminatory ability. His results obtained from data gathered from two able-bodied volunteers are summarised in Table 1.1.

Number of	Hit Rate (%)	
Active Channels	Subject 1	Subject 2
4	94	90
2	92	81

Table 1.1: Vuskovic et al. Hit Rates

Karlik et al. [1994] sought to discriminate between six different arm movements; elbow flexion and extension, wrist supination and flexion, resting and grasping. This was achieved by formulating an autoregressive model of the EMG signal. Two channels of EMG data from the biceps brachii and triceps brachii muscles were split into 12 equal time segments and between two and five autoregressive coefficients were extracted from each of the segments. The resulting feature vectors were then used to train a multilayer perceptron with error back propagation. The multilayer perceptron consisted of between 48 and 120 input neurons, an unspecified number of hidden layer neurons, and 6 output neurons corresponding to the six arm movements. The results obtained using this technique on data obtained from a single able bodied volunteer are summarised in Table 1.2.

Number of AR Coefficients	Hit Rate (%)
48	89
72	94
96	97
120	98

Table 1.2: Karlik et al. Hit Rates

Kelly et al. [1990] sought to discriminate between four arm movements (elbow flexion, elbow extension, wrist pronation, and wrist supination) through the use of a Hopfield network and multilayer perceptron. This approach relies on the Hopfield network's ability to extract features from the raw EMG signals with good discriminatory ability between different contraction patterns. The Hopfield network was used to extract two moving average time series model parameters from a single channel of EMG data. The two features extracted were the first time series parameter and the average signal power. In order to determine the suitability or otherwise of the Hopfield network to provide real-time extraction of these features, a sequential least squares approach was also used to extract the same signal parameters. Unlike the Hopfield network which is an iterative process, the sequential least squares process is an exact algorithm, and traditionally used to perform the parameter extraction. The Hopfield network proved to be between three and five times faster at computing the moving average parameters, depending on the terminating criteria used. The extracted parameters were fed into a multilayer perceptron with error back propagation consisting of two input layer neurons, four hidden layer neurons, and four output layer neurons. Data was processed from several able-bodied volunteers from a pair of electrodes placed over the biceps brachii and triceps brachii muscles. A set of data was also gathered from an amputee volunteer with the electrode position being changed so as to maximise cross talk between the three main muscles in the vicinity. While for all cases a successful classification rate of 100% was reported using the training data, no test data classification rates are mentioned.

Zardoshti-Kermani et al. [1995] performed a thorough study of the effectiveness of a number of features to provide discrimination information. The features examined were the integral of absolute value, the number of zero crossings, the variance (a measure of signal power), and the Willison amplitude. The Willison amplitude is the number of counts for each change in the EMG signal amplitude that exceed a predefined threshold. This feature is

an indicator of muscle contraction level. Other features that were examined included the coefficients of the linear autoregressive time series, the ν-order and log detecto features, and a novel feature known as the EMG histogram.

The ν-order and log detecto approaches to feature extraction are based on a mathematical model of EMG signal generation that is formulated in terms of an ergodic Gaussian process, and an exponential function of muscle contraction force. The EMG histogram feature is calculated by subdividing the raw EMG signal's voltage range into a number of bins, symmetric about the zero voltage axis. The frequency with which the EMG signal enters each of the bins is then calculated. Two techniques were used to judge the effectiveness of each of the above features. The first was the K nearest neighbour feature space classifier in which a data point whose identity is unknown is assigned to the hypercluster represented by the majority of its neighbours. The second technique evaluated cluster separability in the feature space. This is the Davies-Bouldin (DB) cluster separation measure, which is the average of the smallest Euclidean distances between each cluster and each of its neighbouring clusters. The DB measure thus provides an indication of how badly the hyperclusters overlap their nearest neighbours. EMG signals were recorded by two pairs of electrodes, one set placed above the biceps brachii and the other set above the triceps brachii of an above elbow amputee. These signals were recorded as the amputee actuated high, low, and no contraction levels on each muscle, resulting in nine different control states. After feature extraction, a nonlinear transform was used to increase cluster separability in the feature space. This was done so as to provide a visual indicator as to which five features provided the best discriminatory information. These five contraction states were to be used to control an in-house prosthesis with five available movements; elbow flexion, elbow extension, wrist supination, wrist pronation, and no motion. Table 1.3 summarises the range of hit rates achieved by the various features that were trialled.

1.4.2 Frequency Domain-Based Command Discrimination

Boca and Dong [1994] sought to classify between three different levels of contraction of the biceps brachii muscle. This was accomplished by firstly calculating the power spectra of four channels of EMG data. The power spectra represents the continuous distribution of the power of an EMG signal as a function of frequency. The power spectra was then split up into eight harmonics, and percent power and phase angle features were then extracted from each harmonic. The resulting 64 dimensional hyperclusters were then

Feature	Hit Rate Range (%)
Absolute Value	60-75
Zero Crossings	5-25
Variance	0-5
Autoregressive	60-75
ν-Order	75-95
Log Detecto	40-60
Willison Amplitude	75-95
EMG Histogram	95-100

Table 1.3: Zardoshti-Kermani et al. Hit Rates

recursively clustered using a fuzzy c-means algorithm. This resulted in the three hyperclusters each being further refined into three sub-clusters. The application of fuzzy set theory to this problem is appropriate considering the fact that the biological system responsible for producing the EMG signals is very complex and the boundaries between different classes are poorly defined. The 64 dimensional feature vectors were then used as the training data for a multilayer perceptron neural network with error back propagation. The perceptron consisted of 64 input neurons, two hidden layer neurons, and one output neuron. The operating range of the output neuron was split into nine zones, with each zone corresponding to the level of muscle contraction of a particular subcluster in the feature space. Table 1.4 summarises the results obtained by Boca and Dong from an unspecified number of able-bodied volunteers.

Level of Contraction	Hit Rate (%)
Resting	97
Slight	85
Moderate	80
Strong	78

Table 1.4: Boca and Dong's Hit Rates

Ito et al. [1991] have proposed a similar system for discriminating between six different types of arm motion: elbow flexion and extension, wrist supination and flexion, hand opening and hand closing. Data was recorded from four pairs of electrodes located on the forearm near the elbow. Each of the raw EMG signals were then split up into ten equal-length time segments.

These time segments were then individually passed in parallel through three bandpass filters with central frequencies of 70, 160 and 360 Hz, each with a bandwidth of 40 Hz. The 30 EMG records per channel that resulted contain time-frequency information in much the same way as if the raw EMG signals had been processed using a discrete Fourier or discrete wavelet transform. These records were then time averaged, and the results normalised, resulting in a 30×1 feature vector per channel. All of this data processing was performed by a Hopfield neural network. The output from the Hopfield network is then fed as training data to a multilayer perceptron with error back propagation. The multilayer perceptron consisted of 120 input neurons, ten hidden layer neurons, and six output layer neurons, with each output neuron corresponding to one of the six arm movements. Test data from two able-bodied and one amputee volunteer was then fed through the multilayer perceptron with an average successful classification rate of 90%. This was found to be independent of both volunteer, and electrode placement.

Farry et al. [1996] evaluated a number of different schemes suitable for myoelectric control of the Utah/MIT Hand, with a view to space-based remote teleoperation. These control strategies sought to discriminate between two grasps (key and chuck) and three thumb motions (abduction, extension, and flexion). Two single-ended electrodes were used. The first was positioned above the abductor pollicis longus, extensor pollicis brevis muscle pair, and the second was positioned over the flexor pollicis longus muscle. These electrodes were referenced to a third earth electrode that was placed over electrically inactive tissue. The first method used to discriminate between the two grasps involved splitting each channel of raw EMG data into six equal-length time segments. Features corresponding to the mean absolute value, mean absolute value gradient, number of zero crossings, and waveform length were extracted from each segment. Farry reports a successful classification rate of 80% using an unknown classifier. The second technique used to discriminate between the grasps does so on the basis of the spectral properties of each of the data segments. This spectral data is obtained by first applying a Hamming window to each segment of the EMG data. The periodogram (a stationary spectral estimator) is then calculated, resulting in a sparse spectrogram. Four spectral magnitude averages covering the range of 75 to 250 Hz were then extracted from each data segment. The resulting 24 dimensional feature vectors were then presented as training data to a multilayer perceptron with an unspecified number of hidden and output layer neurons. A successful classification rate of 93% was reported when using this technique on a set of test data. Three frequency-based techniques were used to discriminate between the three thumb motions of interest. The first involved the calculation of segment spectral magnitude averages using

the periodogram. The second parameterised the EMG signal as a set of autoregressive model coefficients. The third technique used the relatively new Thomson's multiple window method (MWM), which is a nonparametric spectral estimator. Unlike the single Hamming window approach, the MWM applies a set of windows that are both mutually orthogonal and concentrated in frequency. This results in a spectrum estimator with considerably lower bias and variance than the periodogram. As with the grasp discrimination problem, a multilayer perceptron is used to perform the classification of features, with a 90-95% successful classification rate being achieved.

Kuribayashi et al. [1992] also use the power spectrum of the EMG signal in association with a neural network to perform task discrimination. Seven movements were to be differentiated between; relaxation, wrist pronation, wrist supination, palmar flexion, palmar dorsiflexion, hand opening, and hand closing. The EMG power spectrum were calculated from data obtained from a single able bodied volunteer using a taper window and FFT. The resulting spectral data from four channels is then discretised using an unspecified number of frequencies. The resulting feature vector was then fed to a multilayer perceptron with error back propagation consisting of 100 hidden layer neurons, and three output layer neurons. Kuribayashi reports an average successful classification rate of 81%, with hit rates shown by function in Table 1.5.

Arm Movement	Hit Rate (%)
Relaxation	100
Wrist Pronation	100
Wrist Supination	100
Palmar Flexion	70
Palmar Dorsiflexion	52
Hand Opening	82
Hand Closing	48

Table 1.5: Kuribayashi et al. Hit Rates

1.5 Research Overview

The research presented in this thesis involves the discrimination between six different grasp classes on the basis of features extracted from four channels of EMG data. These four channels of data were recorded from pairs of disc electrodes placed over muscles in the forearm responsible for the movement

of the thumb (extensor pollicis brevis and extensor pollicis brevis), index and middle fingers (extensor communis digitorum), and the little finger (extensor carpi ulnaris). The six grasp classes between which discrimination took place were:

1. **Small Cylinder.** Data for the small cylinder grasp class was obtained as volunteers grasped a cardboard cylinder 30 mm in diameter.

2. **Large Cylinder.** Data for the large cylinder grasp class was obtained as volunteers grasped a metal tin 90 mm in diameter.

3. **Small Sphere.** Data for the small sphere grasp class was obtained as volunteers grasped a tennis ball 50 mm in diameter.

4. **Large Sphere.** Data for the large sphere grasp class was obtained as volunteers grasped a polystyrene ball 150 mm in diameter.

5. **Small Disk.** Data for the small disk grasp class was obtained as volunteers grasped a horizontally mounted metal disc 25 mm in diameter using a thumb-index precision grasp .

6. **Key.** Data for the key grasp class was obtained as volunteers grasped a vertically mounted metal disc 25 mm in diameter using a thumb-index pad palmar grasp.

The observant reader will note that both the grasp classes and the location of the electrodes are identical to those used by Vuskovic [1995, 1996] (see Section 1.4.1). This is due to the fact that in 1996 Prof. Marko Vuskovic paid a visit to the Mechanical Engineering Department of the University of Canterbury. He brought with him a set of his own EMG data and several programs used for processing it. At the time of his visit, the research on which this thesis is based was still very much in its formative stage.

It was decided that the first stage of the research should be to emulate Prof. Vuskovic's discrimination results using the envelope maxima feature. This would provide a baseline of performance against which novel signal parameterisation techniques could be compared. These novel features are the result of one of three parameterisation techniques:

1. **Orthogonal Decomposition.** This method uses a set of orthogonal functions to decompose the EMG signal into a number of orthogonal components. The "amount" of each of the orthogonal functions in the EMG signal is given by the associated coefficient, and it is these coefficients which are used to parameterise the signal.

2. **Global Dynamical Models.** This method seeks to parameterise the dynamical process that produced the EMG signal rather than the EMG signal itself.

3. **EMG Histogram.** This method records the number of times that the EMG signal enters each of a number of voltage bins, and these frequency of entry statistics are used to parameterise the signal. This is a novel application of the concept of the EMG histogram proposed by Zardoshti-Kermani et al. [1995].

1.6 Thesis Outline

The remainder of the thesis is laid out as follows:

Chapter 2. Fundamentals of Anatomy and Physiology. This chapter provides a detailed explanation of the origin and properties of the EMG signal. Such knowledge is necessary given the fact that the EMG signal plays such a pivotal role. The processes responsible for the creation of the EMG signal occur at a molecular level, and it is at this scale that Chapter 2 begins. From here the scale of the investigation rises up through several orders of magnitude before terminating at the level of the gross skeletal muscle. A knowledge of the anatomy of the gross skeletal muscle of the hand, wrist, and forearm is necessary as it determines the placement of the disc electrodes used for recording the EMG signal. Chapter 2 also contains an examination of the anatomy of the skeleton underlying the gross skeletal muscle, which has been included for completeness.

Chapter 3. Experimental Equipment and Methodology. Having had the origin of the EMG signal explained in Chapter 2, Chapter 3 introduces the equipment used to record the EMG signal and the associated experimental methodology. It begins with an explanation of the principles of operation of the disc electrode, and then goes on to discuss the other equipment necessary to record the EMG signals. This includes the leads, amplifier box, and data acquisition card. Chapter 3 concludes with an explanation of the clinical trial procedure used when gathering data from volunteers.

Chapter 4. Mathematical Concepts. This chapter details the mathematical constructs that are used to parameterise the EMG signals. It starts with an examination of the envelope maxima feature before

moving on to discuss the principles behind the technique of orthogonal decomposition and the global dynamical model approach to parameterisation. Chapter 4 then moves on to examine the EMG histogram technique in more detail, before finishing with an examination of the statistical methods used to measure the ability of each of the different parameterisations to provide discrimination information between the grasp classes.

Chapter 5. Experimental Results. This chapter presents the discrimination results obtained from the data collected during the clinical trial. It also contains a detailed discussion of these results and the statistical processes used to calculate them.

Chapter 6. Conclusions. In Chapter 6, the effectiveness of the novel techniques used to discriminate between grasp classes is summarised, and improvements to these techniques are suggested. Also included in Chapter 6 is a list of research areas that bear further investigation in future research.

Chapter 2

Fundamentals of Human Anatomy and Physiology

As discussed in Chapter 1, the research on which this thesis is based consists of parameterising the EMG signal for the purpose of discriminating between the grasp classes. These EMG signals are produced by skeletal muscle tissue as the result of activities taking place at the molecular level. This necessitates not only a study of the anatomy and physiology of the human hand, wrist, and forearm, but also an in-depth examination of the underlying metabolic processes. Skeletal muscle tissue is one of three types of muscle tissue present in the human body:

1. **Skeletal Muscle Tissue.** Most of the muscle mass in the human body is skeletal muscle tissue. It is responsible for skeletal movement, maintaining posture and body position, supporting soft tissues, and guarding the entrances and exits to the digestive and urinary tracts. Most skeletal muscle is under voluntary control.

2. **Smooth Muscle Tissue.** Smooth muscle tissue is found in the respiratory, circulatory, digestive, and reproductive tracts. While smooth muscle is able to contract, it is under the control of the central nervous system and not voluntary control.

3. **Cardiac Muscle Tissue.** Cardiac muscle tissue is found only in the heart. The central nervous system is able to control the rate of contraction of cardiac muscle tissue, but unlike skeletal and smooth muscle tissue, it is not responsible for starting the contraction.

2.1 Skeletal Muscle Tissue

Skeletal muscle tissue is made up of a number of different structures, each at a different scale. Macroscopically, the organisation of skeletal muscle tissue is defined by its connective tissue. There are three layers of connective tissue within each skeletal muscle:

1. **The Epimysium.** Every skeletal muscle is surrounded by the epimysium. Its purpose is to physically separate the muscle from neighbouring tissues and organs.

2. **The Perimysium.** The perimysium divides the skeletal muscle up into a series of compartments called fascicles. It contains collagen and elastic fibres as well as nerves and blood vessels. Each fascicle receives branches of the blood vessels and nerves in the perimysium.

3. **The Endomysium.** Each fascicle contains a number of individual muscle fibres. These are separated within the fascicle by the endomysium, which also serves to tie adjacent muscle fibres together.

These three layers of connective tissue serve to compartmentalise successively smaller functional units. This fact is illustrated by Figure 2.1. The collagen fibres of the endomysium and perimysium are interwoven, and join with those of the epimysium at the end of the muscle to form a bundle known as a tendon.

Skeletal muscle tissue is made up of thousands of individual skeletal muscle fibres, or skeletal muscle cells. Each muscle fibre is formed from groups of embryonic cells called myoblasts which have fused together. This formation process results in muscle fibres having multiple nuclei and being significantly larger than "typical" cells. The genes contained within these multiple nuclei direct the production of enzymes and structural proteins required for normal muscle function. The sarcolemma (cell membrane) surrounds the sarcoplasm (cell cytoplasm) of each muscle fibre. The sarcolemma has a characteristic transmembrane potential that is caused by the unequal distribution of positive and negative ions across it. The unequal distribution of ions is the result of a dynamic equilibrium existing between the various mechanisms that transport substances across the sarcolemma. The transmembrane potential plays an important role in the generation of the EMG signal. It is through changes in the transmembrane potential that the specialised activities of certain cells are triggered. In the case of skeletal muscle fibres, that specialised activity is contraction.

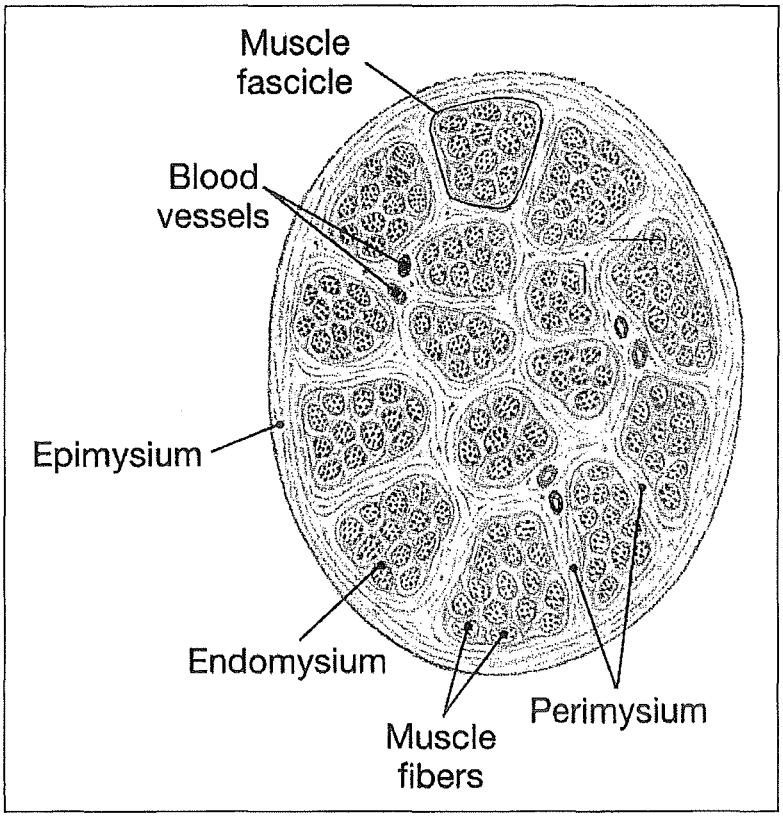

Figure 2.1: The Three Layers of Connective Tissue Within Skeletal Muscle. [Martini 1998, p. 278].

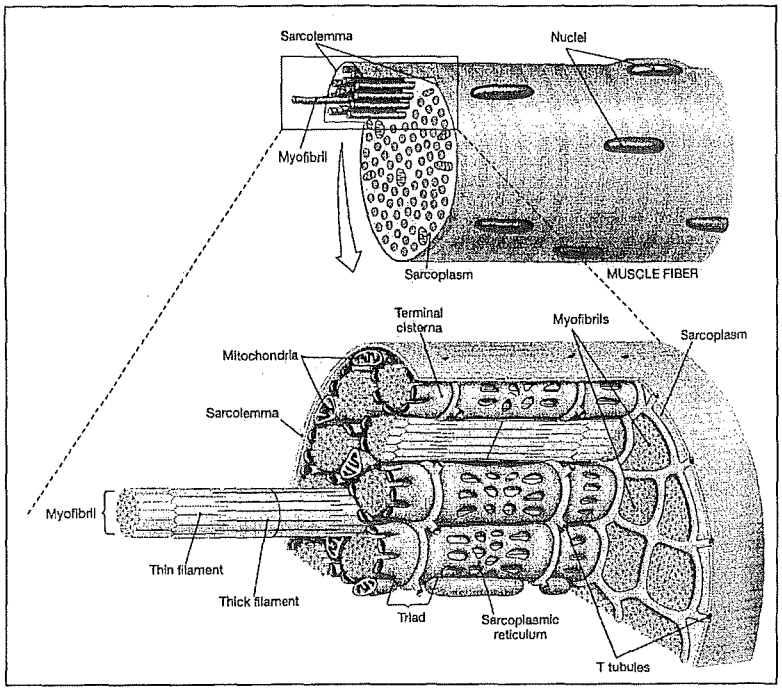

Figure 2.2: The Structure of a Skeletal Muscle Fibre. [Martini 1998, p. 280].

2.1.1 Anatomy of Skeletal Muscle Fibres

Skeletal muscle fibres are very large, but all regions of the muscle fibre must contract at the same time. Therefore the signal to contract must be distributed quickly throughout the interior of the muscle fibre. This signal conduction occurs at the transverse tubules. The transverse tubules are narrow tubes that are continuous with the sarcolemma and extend into the sarcoplasm at right angles to the membrane surface (Figure 2.2). Filled with extracellular fluid, they form a network through which electrical impulses can travel deep into the skeletal muscle fibre. These electrical impulses, or action potentials, are the trigger for muscle contraction. Inside the muscle fibre the branches of the transverse tubules encircle cylindrical structures called myofibrils (see Figure 2.2). Myofibrils are composed of bundles of myofilaments. These myofilaments are protein filaments composed primarily of actin or myosin. The hundreds of myofibrils in each muscle fibre can actively shorten and are responsible for muscle fibre contraction. At each end of the muscle fibre the myofibrils are firmly attached to the sarcolemma. The sarcolemma itself is firmly bound to collagen fibres that extend into the tendon of the skeletal muscle and thus attach it to the bone. Wherever a transverse tubule encircles a myofibril, the tubule makes close contact with the mem-

branes of the sarcoplasmic reticulum. On either side of a transverse tubule, the tubules of the sarcoplasmic reticulum enlarge, fuse and form expanded chambers called terminal cisternae. The combination of a pair of terminal cisternae plus a transverse tubule is known as a triad.

A myofibril consists of approximately 10 000 functional units called sarcomeres that are joined end to end. Sarcomeres are the smallest functional units of the muscle fibre. Interactions between the thick and thin filaments of sarcomeres are responsible for muscle contraction.

2.1.2 Microanatomy of Skeletal Muscle Fibres

A thick filament consists of approximately 500 myosin protein molecules bound to a titin protein core. Each of these myosin molecules consists of two distinct parts: the head and the tail. The tail of the myosin molecule is bound to the other myosin molecules within the thick filament (Figure 2.3). The free head of the myosin molecule consists of two globular subunits. It is attached by a hinge to the tail of the myosin molecule, and points out of the thick filament towards the nearest thin filament. During contraction, the myosin heads interact with active sites on the thin filaments, and are thus also referred to as cross bridges. Thin filaments contain three proteins:

1. **F Actin.** This is a twisted strand of several hundred globular G actin molecules arranged in a double helix. Each G actin molecule contains an active site that is able to bind to the head of a thick filament. Under resting conditions this is prevented from happening by the actions of the troponin and tropomyosin molecules.

2. **Tropomyosin.** Strands of the molecule tropomyosin cover the active sites on the thin filaments and prevent interaction between the G actin of the thin filament and the myosin heads of the thick filaments. Each molecule of tropomyosin covers seven active sites on the F actin strand and is bound to one molecule of troponin midway along its length.

3. **Troponin.** A troponin molecule consists of three globular subunits: one subunit binds to tropomyosin, locking them together; a second subunit binds to G actin, holding the troponin-tropomyosin complex in position on the F actin strand; and the third subunit has a receptor that is able to bind a calcium ion (Ca^{2+}). In a resting muscle the intracellular Ca^{2+} concentrations are very low, and so the binding site is empty.

These three proteins are shown in Figure 2.3. Contraction cannot occur unless there is a change in the position of the troponin-tropomyosin complex

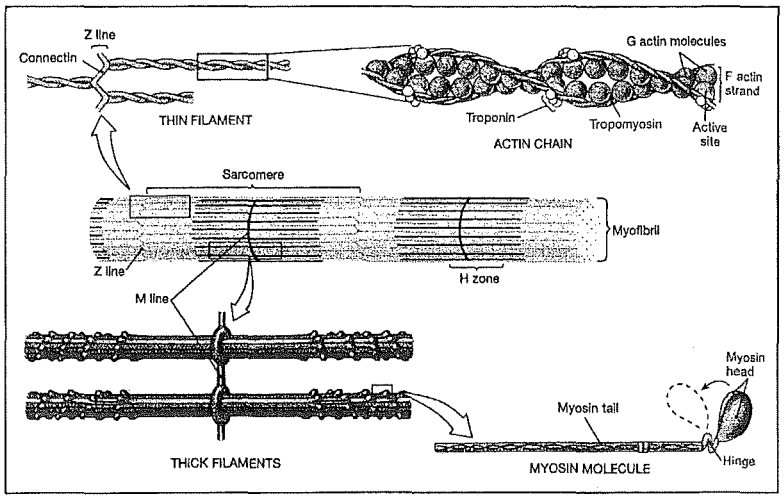

Figure 2.3: The Thick and Thin Myofilaments. [Martini 1998, p. 284].

that exposes the active sites on the F actin. This change in position is brought about when calcium ions bind to receptors on the troponin molecules.

2.2 Contraction of Skeletal Muscle Tissue

Skeletal muscle fibres are solely under the control of the central nervous system. Communication between the central nervous system and a skeletal muscle fibre occurs at a specialised intercellular connection known as a neuromuscular junction.

2.2.1 The Neuromuscular Junction

Each skeletal muscle fibre is controlled by a neuron at a single neuromuscular junction midway along the fibre's length. A single axon branches out in the perimysium to form a number of fine branches. Each of these branches ends in a synaptic terminal on a different muscle fibre. The cytoplasm of the synaptic terminal contains (among other things) a number of vesicles filled with molecules of acetylcholine (ACh). ACh is a neurotransmitter able to alter the membrane properties of the sarcolemma and thus trigger a contraction of the muscle fibre. Between the synaptic terminal of the neuron and the surface of the opposing sarcolemma is a small gap called the synaptic cleft. The region of the sarcolemma immediately beneath the synaptic terminal is known as the motor end plate. The motor end plate surface contains membrane receptors that are able to bind to ACh. The motor end plate contains

a number of deep creases whose purpose is to increase surface area and thus also the number of available ACh receptors. The synaptic cleft and motor end plate also contain the enzyme acetylcholinesterase (AChE) which breaks down the neurotransmitter ACh. The stimulation of a muscle fibre by a neuron takes place in several steps:

1. **Arrival of an Action Potential** The stimulus for ACh release is the arrival of an electrical impulse or action potential at the synaptic terminal.

2. **Release of ACh** When the action potential reaches the synaptic terminal permeability changes in the membrane of the synaptic terminal trigger the release of ACh into the synaptic cleft.

3. **ACh Binding** ACh molecules diffuse across the synaptic cleft and bind to the ACh receptors on the motor end plate's surface. This binding changes the permeability of the sarcolemma to sodium ions. Sodium ions rush into the sarcoplasm until AChE removes the ACh from the receptors.

4. **Appearance of an Action Potential in the Sarcolemma** The sudden inrush of sodium ions results in the generation of an action potential in the sarcolemma at the edges of the motor end plate. This electrical impulse sweeps across the entire membrane surface and travels along the transverse tubules.

5. **Return to the Initial State** Even before the action potential has spread across the entire membrane, the ACh has been broken down by AChE. Some of these breakdown products will be absorbed by the synaptic terminal and used to resynthesize ACh for subsequent release.

2.2.2 The Contraction Cycle

The link between the generation of an action potential in the sarcolemma and the start of a muscle contraction is called excitation-contraction coupling. This coupling occurs at the triads. When an action potential reaches a triad it triggers the release of Ca^{2+} from the terminal cisternae of the sarcoplasmic reticulum. Because the terminal cisternae are situated directly over the zones of overlap between the thick and thin myofilaments, the effect of the release of Ca^{2+} is almost instantaneous. Contraction begins, and occurs in a number of steps:

1. **Active Site Exposure.** The Ca^{2+} entering the sarcoplasm binds to the receptor on troponin. This weakens the bond between troponin and G actin. The troponin molecule then changes position and pulls the tropomyosin molecule away from the active sites on the F actin.

2. **Cross Bridge Attachment.** When the active sites on the thin myofilament are exposed, the heads of the thick molecules bind to them forming cross bridges.

3. **Pivoting.** In the resting sarcomere, the heads of all of the myosin molecules that make up the thick myofilaments point away from the M line. In this position they are primed. In order to prime the heads, energy is required which was obtained from the breakdown of ATP into ADP and a phosphate group. In the primed position, both the ADP and the phosphate group are still bound to the myosin head. Once the cross bridge have formed, this stored energy is released as the myosin heads pivot towards the M line. This is called the power stroke. After this occurs the ADP and phosphate group are released.

4. **Cross Bridge Detachment.** When ATP binds to the myosin head, the link between the active site on the actin molecule and the myosin head is broken.

5. **Myosin Reactivation.** Myosin reactivation occurs when the free myosin head splits the ATP into ADP and a phosphate group.

Providing that calcium levels within the sarcomere remain elevated and there is sufficient ATP available then this cycle will occur approximately five times a second. The amount of tension produced in a muscle fibre undergoing such a contraction is dependant on its initial starting length. This determines the extent to which the thick and thin myofilaments overlap and thus the number of heads of the myosin molecules that are able to form cross bridges. The greater the overlap, the larger the number of myosin heads that are able to from cross bridges and contribute to the overall tension of the muscle fibre. The number of myosin heads in the zone of overlap that form active bridges is not able to be controlled. As such the muscle fibre obeys the so-called all or nothing principle, whereby either the muscle fibre is producing no tension at all (resting) or is producing all the tension that it is capable of. The amount of tension produced by a skeletal muscle is dependant on the number of muscle fibres stimulated and the frequency of stimulation.

A twitch is a single stimulus-contraction-relaxation sequence in a muscle fibre. A single twitch can be divided into 3 stages:

1. **Latent Period.** The latent period begins at stimulation. During this period the action potential sweeps across the sarcolemma. No tension is produced during this phase however as the contraction cycle has yet to begin. The latent period lasts about 2 msec.

2. **Contraction Phase.** In the contraction phase tension rises to a peak. The contraction phase ends approximately 15 msec after stimulation.

3. **Relaxation Phase.** The relaxation phase continues for another 25 msec after this. During the relaxation phase, Ca^{2+} levels are falling, active sites are being covered by tropomyosin, and the number of active cross bridges is falling.

A single twitch cannot be used to perform any useful task. Normal activities are achieved through sustained muscle contraction. If only one action potential were to sweep the sarcolemma, then tension production would be completed 17 msec after the contraction phase began. This is caused by the sarcoplasmic reticulum scavenging Ca^{2+} out of the cytosol and back into the terminal cisternae. The creation of a sustained muscle contraction therefore relies on the arrival of a number of action potentials at the motor end plate in short succession. If a second stimulus arrives at the sarcomere before the the relaxation phase has ended, then a second more powerful contraction occurs. The addition of one twitch to another in this manner is called wave summation. The frequency of stimulation necessary to produce wave summation depends on the duration of the twitch. For instance, if a single twitch lasts 20 msec, stimulation at a frequency of less than 50 Hertz will result in a series of individual twitches. Stimulation at a frequency higher than 50 Hertz will produce wave summation. If stimulation continues and the muscle is never allowed to relax completely then tension will rise to a peak. If the muscle is stimulated at a still higher rate such that the relaxation phase is eliminated completely, then the muscle achieves a state of complete tetanus. During complete tetanus, the action potentials are arriving so fast that the sarcoplasmic reticulum does not have time to scavenge Ca^{2+} out of the cytoplasm. The high Ca^{2+} concentration in the cytoplasm prolongs the state of contraction, making it continuous. Virtually all normal muscular contractions involve the complete tetanus of the participating muscle fibres.

When a muscle fibre is in a state of complete tetanus, its ATP reserves are very quickly depleted. They are able to be kept up at usable levels for approximately 17 seconds through the conversion of ADP back to ATP. This is achieved through the use of the creatine phosphate (CP) molecule. A resting muscle fibre produces more ATP than it needs. The muscle fibre uses this excess energy to convert creatine into creatine phosphate. These CP

reserves are then able to be used to convert ADP back into ATP, turning back into creatine in the process.

Skeletal muscle fibres are connected together within a fascicle into groups called motor units. A motor unit may contain between five and several thousand individual muscle fibres depending on the particular muscle. Each motor unit is controlled by a single nerve. As a movement is started, the smallest motor units within a muscle begin to exert a force. Over time larger and more powerful motor units add to the force produced through a process called recruitment.

During sustained tetanic contraction, motor units within a muscle are activated on a rotating basis. Some will be resting and recovering while others will be actively contracting. This process is called asynchronous motor unit contraction. As a result, when skeletal muscle is contracted for sustained periods of time, it produces less than maximal tension.

The EMG signal as recorded at the dermal layer is the spatial summation of the electrical activity of all of the active motor units within the local vicinity.

2.3 Skeletal Anatomy

Having detailed the various mechanisms which contribute to the ability of skeletal muscle to contract, it is now time to examine the skeletal structure of the human hand, wrist, arm, and upper limb. Whilst the skeletal muscles are responsible for generating the forces by which the skeleton is animated, it is the configuration of the skeleton that determines what range of motions are available.

2.3.1 The Humerus

The upper arm contains one bone: the humerus. The head of the humerus is attached to the glenoid cavity of the scapula through a ball-and-socket joint. This articulation provides the shoulder with the greatest range of movement of any joint in the body. Near the distal end of the humerus, the shaft expands to form the medial and lateral epicondyles. This broad triangle of bone provides greater room for muscle attachment than would otherwise be possible. At the condyle, the trochlea of the humerus articulates with the bones of the forearm: the radius and ulna.

2.3.2 The Ulna

The ulna and radius are parallel bones that support the forearm. The olecranon process of the ulna forms the point of the elbow. The trochlea notch of the ulna articulates with the trochlea of the humerus at the elbow joint. Near the wrist the ulnar shaft narrows before ending in a disc shaped head that bears a short styloid process. The articular cartilage attaches to the styloid process, separating the ulnar head from the bones of the wrist.

2.3.3 The Radius

The radius is the lateral bone of the forearm. The disc shaped head of the radius articulates with the capitulum of the humerus. The radial tuberosity marks the attachment site of the biceps brachii which is responsible for flexing the forearm. The distal end of the radius articulates with the bones of the wrist. The medial surface of the distal end of the radius articulates with the ulnar head at the ulnar notch. The articulation of the radius and ulna enables the ulnar notch of the radius to roll across the rounded surface of the ulnar head. This movement is called pronation, and the reverse movement that returns the radius and ulna to their anatomical positions is called supination.

2.3.4 The Carpal Bones

The eight carpal bones on the wrist form two rows. The four proximal carpal bones are the scaphoid, lunate, triangular, and pisiform bones. The four distal bones are the trapezium, trapezoid, capitate, and hamate bones. The carpal bones articulate with one another at joints that permit limited sliding and twisting. Ligaments connect the carpal cones and help stabilise the wrist joint.

2.3.5 The Hand

Five metacarpal bones articulate with the distal carpal bones of the wrist and support the hand(Figure 2.4). Distally, the metacarpal bones articulate with proximal finger bones. There are 14 phalanges or finger bones in the hand. The thumb has two phalanges (proximal and distal), while each of the other fingers has three (proximal, middle, and distal).

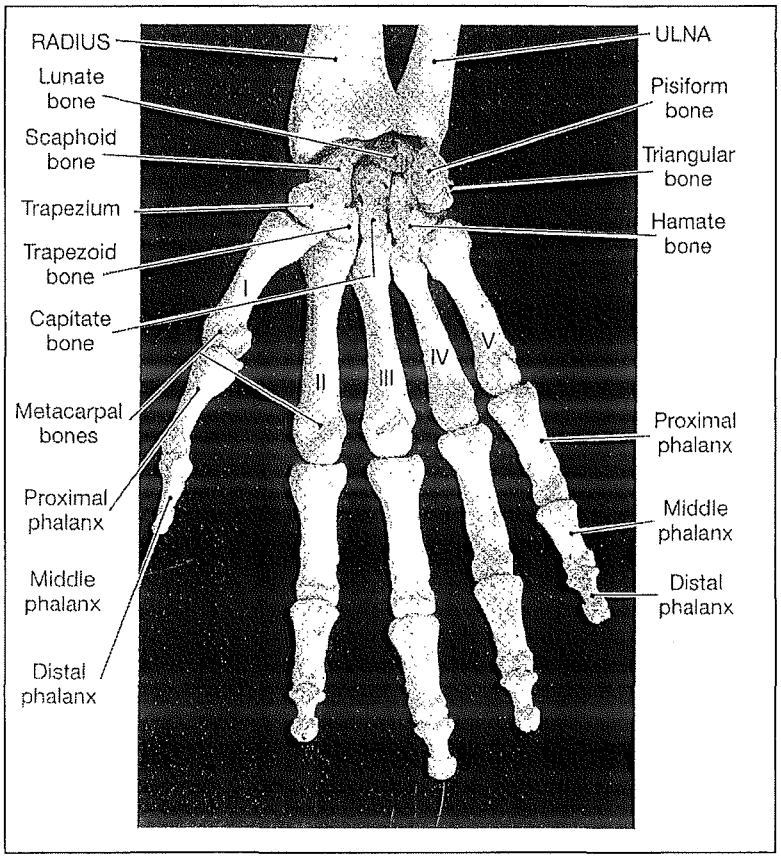

Figure 2.4: The Bones of the Wrist and Hand. [Martini 1998, p. 239].

2.4 Muscle Anatomy

The following is a summary of the muscles that are responsible for the movement of the arm, wrist, and hand.

2.4.1 Muscles that Move the Arm

Abduction of the arm is performed by the deltoid, with assistance from the supraspinatus at the beginning of the movement. The subscapularis and teres major rotate the arm medially, while the infraspinatus and teres minor perform lateral rotation. The coracobrachialis is responsible for adducting the humerus. The pectoralis major flexes the arm at the shoulder joint, and the latissimus dorsi extends it.

2.4.2 Muscles that Move the Forearm and Hand

Most of the muscles that move the forearm and hand originate within the forearm. This is not the case with the biceps brachii and triceps brachii however, which insert from the arm. Contraction of the biceps brachii flexes the elbow and supinates the forearm. Contraction of triceps brachii extends the elbow. The brachialis and brachioradialis flex the elbow and are opposed by the anoconeus and the triceps brachii respectively. The flexor carpi radialis flexes and abducts the wrist, while the flexor carpi ulnaris flexes and adducts. The flexor carpi radialis and flexor carpi ulnaris are aided in the flexion of the wrist by the palmaris longus. The extensor carpi radialis produces extension and abduction of the wrist, while the **extensor carpi ulnaris** produces extension and abduction. The pronator teres and the supinator arise on both the humerus and forearm. They rotate the radius without either flexing or extending the elbow. The pronator quadratus arises on the ulna and assist the pronator teres in opposing the actions of the supinator or biceps brachii.

2.4.3 Muscles that Move the Fingers

Several superficial and deep muscles of the forearm flex and extend the finger joints. The muscles stop before reaching the wrist however, with only their tendons crossing over into the hand. The tendons pass through tendon sheaths whose purpose is to reduce friction and to keep the tendons lubricated. This provides the hand and wrist with maximum mobility, while at the same time providing muscle forces far in excess of those which the intrinsic muscles of the hand are able to generate. These so-called extrinsic muscles provide strength and crude control of the hand and fingers. Fine

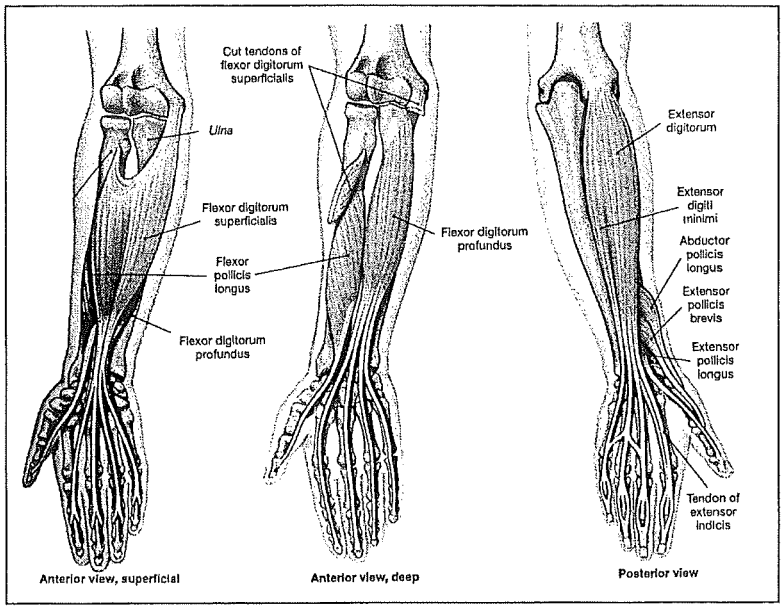

Figure 2.5: Muscles that Move the Fingers and Hand. [Martini 1998, p. 349].

control of the hand is performed by the intrinsic muscles which originate on the carpal and metacarpal bones. No muscles originate on the phalanges, and only tendons extend across the distal joints of the fingers. All of the muscles that move the fingers and hand are shown in Figure 2.5, and a summary of their actions is given in Table 2.1.

Muscle	Action
Abductor Pollicis Longus	Abducts thumb
Extensor Digitorum	Extends fingers and hand
Extensor Pollicis Brevis	Extends thumb, abducts hand
Extensor Pollicis Longus	Extends thumb, abducts hand
Extensor Indicis	Extends and adducts little finger
Extensor Digiti Minimi	Extends little finger
Flexor Digitorum Superficialis	Flexes fingers
Flexor Digitorum Profundus	Flexes distal phalanges
Flexor Pollicis Longus	Flexes thumb

Table 2.1: Muscles that Move the Fingers and Hand. Muscles from which EMG Signals are Recorded are Shown in Bold. [Martini 1998, p. 348].

Chapter 3

Experimental Equipment and Methodology

Having detailed the processes by which the EMG signal is generated within skeletal muscle, an examination of the equipment and procedures used to record them is now appropriate.

EMG signals from selected muscles in the forearm were picked up by pairs of disc electrodes. These very weak signals were transmitted by shielded cables to an amplifier box. The resulting boosted signals were then recorded directly to the hard drive of a personal computer after having been sampled by an A/D card. Each of these components will now be examined in detail.

3.1 Disc Electrodes

Disc electrodes are able to pick up the weak EMG signals given off by skeletal muscle tissue. This is achieved by establishing electrical contact with the skin over the muscle of interest with the aid of an electrolytic gel or paste. The combination of the electrode and electrolyte form a chemical half-cell in which ion exchange is able to occur. In the case of the silver-silver chloride electrodes used in this research, silver ions from the electrodes migrate into the electrolyte, while at the same time ions from the electrolyte combine with the silver-silver chloride electrode. This ion exchange results in the establishment of a charge gradient across the electrode-electrolyte interface. The resulting voltage is known as the half-cell potential. The reaction occurring at the interface of the electrodes [Geddes and Parker 1975] used during this research is given by:

$$\mathrm{Ag} + \mathrm{Cl}^- \rightleftharpoons \mathrm{AgCl} + \mathrm{e}^- \tag{3.1}$$

Note the use of the double arrow in Equation (3.1), indicating that the reaction is proceeding in both directions. When moving from left to right, a half-cell potential of +0.2225 volts is observed, whereas when moving from right to left a half cell potential of -0.2225 volts is observed. It is logical to assume then that two such electrodes placed in a conducting solution will produce a net voltage of zero, given that the anode will be producing a half-cell potential of equal magnitude but of opposite sign to the cathode. This is however not the case. A small 2.5 mV potential exists between the two electrodes. The fluctuating noise current resulting from this potential can however be greatly reduced by coating the electrodes in a uniform film of material. Hence the electrodes used during this research are coated in a thin layer of silver chloride.

In addition to developing a half-cell potential, each electrode has an impedance associated with it. Through the impedance of both electrodes and the rest of the circuit flows a current derived from the bioelectric event. As the input impedance of the amplifier box used to boost the signal strength is high, the voltage drop caused by the electrode impedance is negligible. This is important as the goal is to measure the potential difference of the bioelectric event without drawing current.

While disc electrodes are susceptible to artifacts caused by electrical instability, the main cause of signal artifact is mechanical disturbance of the electrode. Because the electrolyte-electrode interface is a region of charge gradient, any disturbance results in a fluctuating voltage. While these fluctuations are small in magnitude, they are quite large in relation to the bioelectric event of interest. Thus it is important that the interface between the skin, electrolyte, and electrode be kept free of relative movement. This was accomplished by firstly securing the disc electrode to the surface of the skin with circular self-adhesive plasters. These were then held securely in place by strips of strong adhesive tape.

3.2 Leads

The leads used to connect the disc electrodes to the amplifier box consisted of two individually insulated seven strand copper conductors. These were both surrounded by a spiral woven copper strand shield and covered in a tough outer plastic sheath.

At the electrode end of the lead, 55 mm of the outer layer and shield were removed. Ten mm of insulation were then carefully stripped from one of the conductors. The exposed copper strands were then wound around one another to form a single bundle to improve their strength and manageability.

This bundle was then folded in half length-ways and coated with solder to increase its surface area. Over the length of the insulator between the copper-solder terminal and the outer layer and copper shield was placed a length of heat shrink. This was done to improve the mechanical strength of the thin signal-carrying conductor. A stainless steel sleeve of 3.2 mm in diameter and 15 mm long was then placed over the terminal and crimped into place. The sleeve was positioned in such a way that five millimetres of the sleeve projected behind the terminal and was crimped to the double-insulated layer itself. This ensured that any tensile forces applied to the lead would be born by the insulation rather than the fragile terminal itself. A length of heat shrink was then applied to the sleeve, starting five millimetres from its end and continuing five millimetres past it up the insulated conductor. This was done to prevent the stainless steel sleeve from making electrical contact with the volunteer and thus picking up EMG activity, and also to keep foreign debris out of the sleeve.

The same procedure was then applied to the other conductor after it had been cut 25 mm short. This was done so as to guarantee a 25 mm gap between the electrodes when attached to the arm of a volunteer. The ends of the stainless steel sleeves into which the disc electrodes fit were crimped slightly so as to make them elliptical in cross section. This ensured that while the electrode could be easily slipped in to the end of the sleeve, mould flash on the shaft of the electrode wedged it in place when it was turned through 90 degrees.

Attached to the amplifier box end of the leads is a 3.5 mm three-pole stereo plug. This provides connection points to the amplifier box for the two signal-carrying conductors and the copper shield. The length of the leads was kept as short as possible to reduce artifacts caused by mechanical movement of the leads. In addition to the four signal-carrying leads, a fifth active-earth lead is also used when gathering EMG data. It consists of a single solid copper core onto which is crimped a stainless steel sleeve at one end and a banana plug at the other. When gathering data, it is placed on the upper limb opposite to those from which the EMG signals are being recorded.

3.3 Amplifier

The purpose of the amplifier box is to boost the EMG signal amplitude into the range where it can be sampled by the A/D card. The amplifier box applies a gain of 6000 to the four incoming channels of data as well as an analog fourth-order Butterworth filter with a cut-off frequency of 2500 Hz . The filter's purpose is to prevent aliasing caused by the inclusion of very

high frequency signal components. The amplifier box was designed and built by technicians from the University of Canterbury's Mechanical Engineering Department. Physically it consists of a PCB mainboard into which up to four PCB daughterboards may be slotted. At present the amplifier box contains two identical daughterboards, each of which is responsible for processing the EMG data from two channels. On each of the two daughterboards is mounted a single Burr Brown dual channel instrumentation amplifier. These very high accuracy amplifiers were selected on the basis of their low offset voltage, drift, and high common mode rejection properties. In addition to filtering the incoming signal and boosting its power, the daughterboards are responsible maintaining the offset voltage of the incoming EMG signal at zero. This is done by driving the active ground at the mean value of the four EMG signals. In addition, the amplifier box is responsible for driving the shields on each of the leads so as to remove any common mode voltages that may be present. This is done to eliminate any capacitative effects that may be present due to the person to whom the leads are attached acting as a bluff aerial. The processed EMG signals are sent to the A/D card via a 37 pin ribbon cable.

3.4 Data Acquisition Card

The data acquisition card converts the analog EMG signal from the amplifier box into digital data which can then be saved and manipulated by a personal computer. The particular card used was an Adaptec 818-HG data acquisition card with which data was sampled at 5000 Hertz from each channel by a standard 12-bit A/D converter. The 818-HG differs from normal data acquisition (DAQ) cards in that it contains a FIFO (first-in first-out) buffer.

Before the advent of the FIFO buffer, all data acquisition had to be done on a computer running the Microsoft DOS operating system. This is due to the fact that Windows 95, like its predecessor Windows 3.1, is able to assume control of the system whenever it sees fit. This so-called Windows latency problem resulted in the possibility of data being lost if Windows assumed control just as the data was about to be passed to main memory and the signal for a new A/D conversion to begin had been given.

The inclusion of the FIFO buffer solves the Windows latency problem simply and effectively. During the course of normal operation, data from the A/D conversion is put onto the top of the FIFO buffer, which is in essence a 1 kiloword stack. Once all eight channels have been sampled, the data is moved from the FIFO buffer into main memory. If however Windows assumes control at the point where the data would have been transferred,

all that happens is that all the data converted while Windows has control is added to the stack. This is then all transferred to main memory in one block.

3.5 Data Acquisition Program

The program that performed the data acquisition, along with every program used during the course of this research, was written for use in the Matlab™ (matrix laboratory) matrix manipulation program. Such programs are given the suffix ".m" in reference to the M-file scripting language in which they were written. The data acquisition program, DAQ.m, works as follows. Data acquisition begins when the user presses a key when prompted to do so, and continues for five seconds. The resulting data is notch filtered by a second order Butterworth filter with corner frequencies of 49 and 51 Hz. This is done to remove any line noise present in the data. The filtered data from the four active channels is then presented to the user who then has the option of saving it to file or discarding it.

3.6 Experimental Methodology

The results that will be presented further on in this thesis are based on data that was gathered from ten volunteers. The procedure followed when gathering this data is as follows.

Each volunteer had their dominant arm cleaned with alcohol before the location of the four muscles of interest was marked with an indelible non-toxic pen. These muscle locations were established by having the volunteer move the digits of the hand to which the muscles of interest connect, and feeling around in the known general location of these muscles using the tips of the index and middle fingers until the greatest (and most independent) movement was felt. Two locations 25 mm apart were then marked. The muscles that were targeted were the extensor pollicis brevis and extensor pollicis brevis (responsible for moving the thumb), the extensor communis digitorum (middle finger and index finger), and the extensor carpi ulnaris (little finger).

The two disc electrodes from each lead were then secured in place over the marked locations using the hole punched plasters. Once all four leads were in place, strips of adhesive tape were stretched over the electrodes to ensure that no movement relative to the skin's surface was possible. All four leads were then fed up past the elbow of the volunteer where they were held in place

by a piece of tape. This was done to ensure that there was no chance of the electrode positions being altered by the leads being pulled. Once the earth lead had been secured onto the opposite arm, all the disc electrodes were filled with electrolytic gel using a 10 cc syringe and blunted hypodermic needle. The volunteer was then asked to move the digits of the wired hand and data was recorded to ensure that a good signal-to-noise ratio was being obtained from each of the four leads. Once this had been established, the volunteer was briefed as to how the data collection session would be run. They were informed of the necessity to reach out to grasp each object at the same speed and location, and to envelop the object of interest but not to actually touch it. Repeatability of grasp speed and location is important so that all the data records for a particular object are as similar as possible. Likewise it is important that the object not actually be grasped, as this introduces EMG activity from the isometric contraction that does not contribute any significant discriminatory information, and due to its magnitude swamps the EMG signal of the pregrasp motion.

Data was then recorded as the volunteer grasped each one of the six objects ten times in a row. Each grasp was initiated by the experimenter giving a verbal command. The volunteer started each grasp with their hands resting comfortably on their thighs with their hands in a neutral position. They were told to totally relax both the arm and hand in order to minimise EMG activity in the limb as the grasp run began. After having enveloped the object the hand and arm was returned to the starting position before being relaxed once more. A five second rest period between grasps and a one minute rest between objects ensured that muscle fatigue and its associated spectral shifting of the EMG signal was avoided. This cycle of 60 grasps was repeated ten times to give a total of 100 grasp records for each of the six objects per volunteer. The total data per person came to 997 MB. Figures 3.1–3.6 show samples of the raw EMG signals that were recorded from each of the four sets of electrodes.

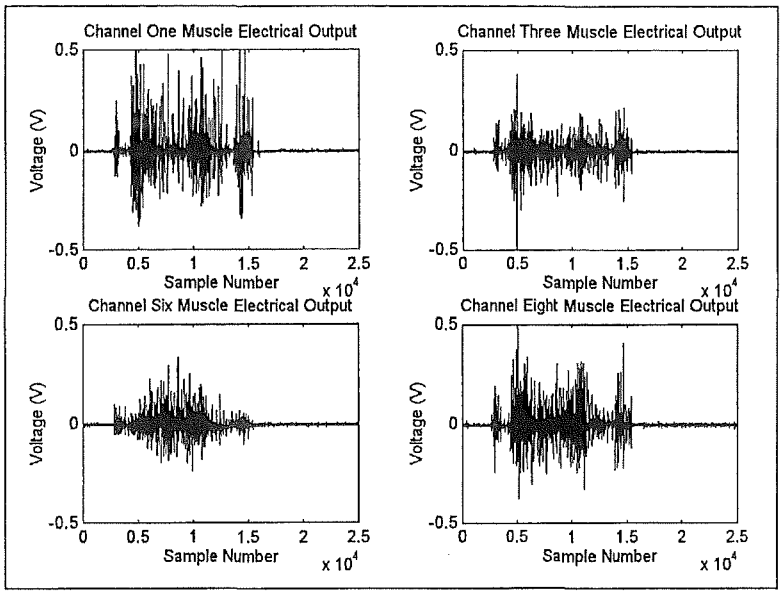

Figure 3.1: Small Cylinder Grasp Sample Data.

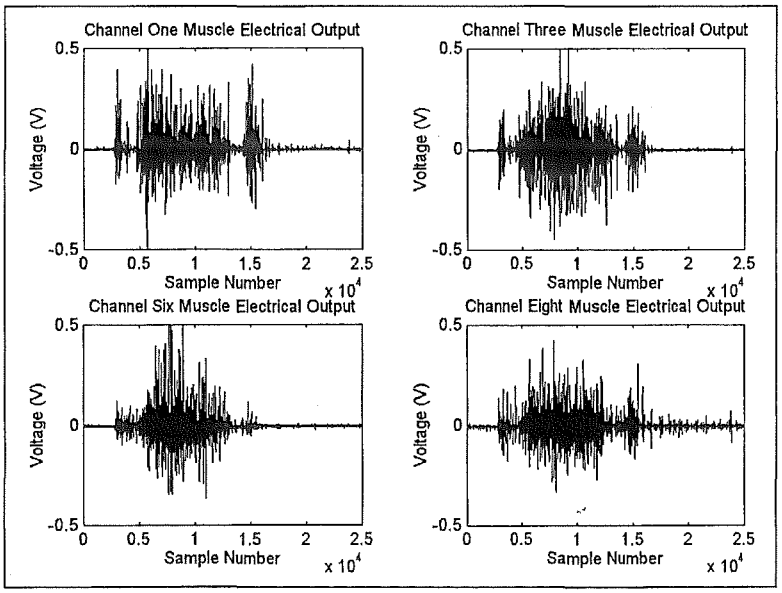

Figure 3.2: Large Cylinder Grasp Sample Data.

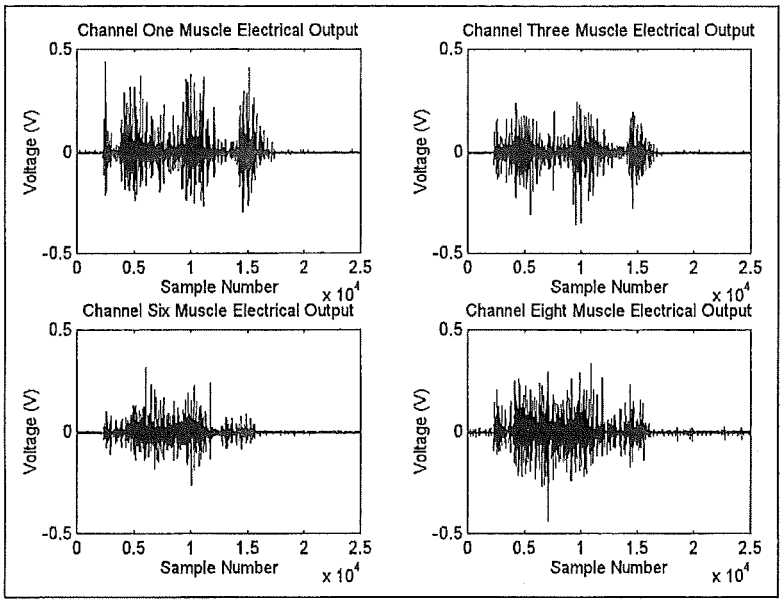

Figure 3.3: Small Sphere Grasp Sample Data.

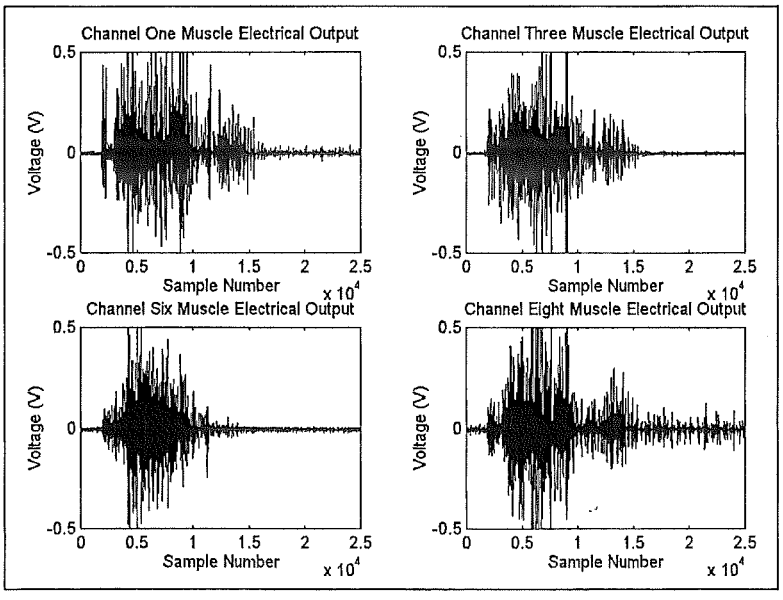

Figure 3.4: Large Sphere Grasp Sample Data.

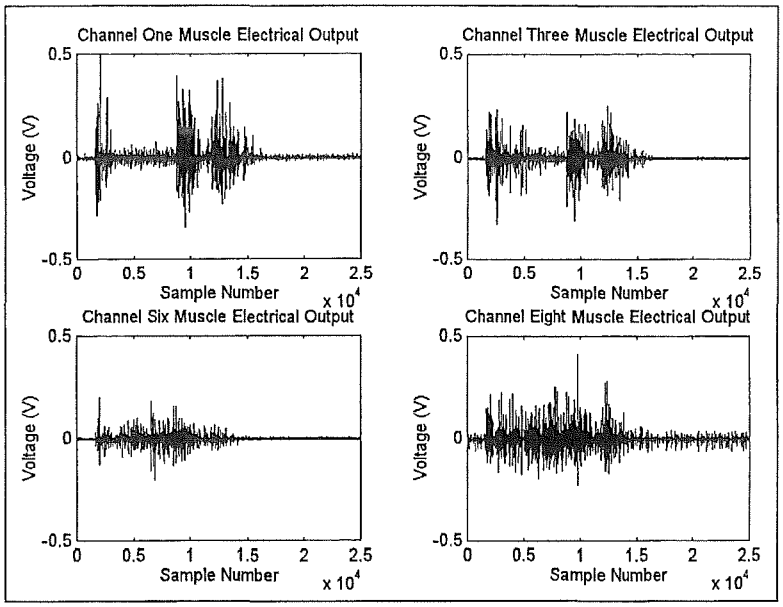

Figure 3.5: Small Disk Grasp Sample Data.

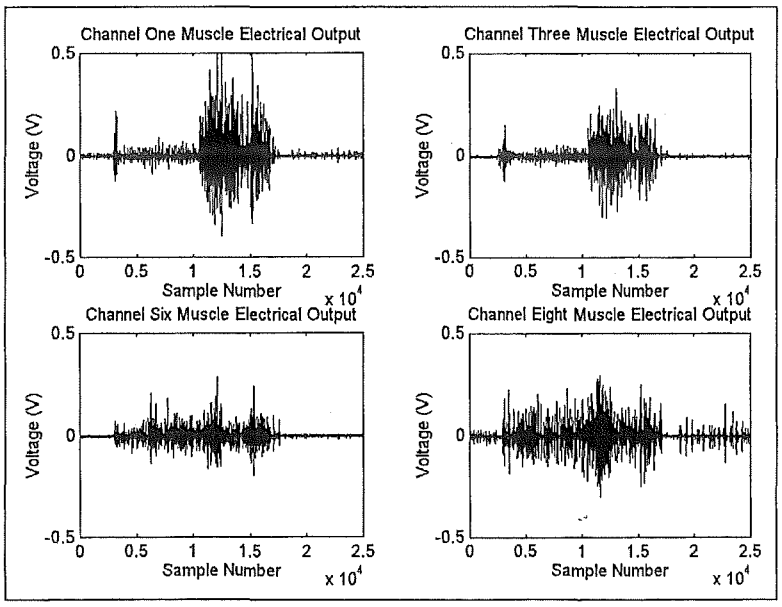

Figure 3.6: Key Grasp Sample Data.

Chapter 4

Mathematical Concepts

In order to appreciate just what it is that differentiates one parameterisation technique from another, it is necessary to examine their underlying mathematical basis.

A *grasp* can be thought of as a collection of n time series $f_i(t), i = 1, \ldots, n$ with each time series representing the amplitude of an EMG signal collected simultaneously from one of n locations on the forearm. Physically, each of these n time series corresponds to a channel of recorded data for which $0 \leq t \leq T$, where T is the duration of a typical burst.

Assuming that $n = 4$ channels, $T = 2$ seconds, and that each channel is sampled at 5 kHz, then a single grasp is characterised by 40 000 signal values. This is far too many points to allow real-time grasp discrimination to take place. The solution to this problem is to parameterise each time series in terms of a small number of *features*.

Several different methods of extracting features from the EMG signals were tested during the course of this research. The first of these involved the parameterisation of the time series $f_i(t), i = 1, \ldots, n$, by their envelope maxima. This well known time domain feature was chosen to allow comparison of results with those of others working in the field (see Section 1.4), and also to provide a baseline measure of performance to which the novel parameterisation techniques developed during the course of this research could be measured. The second parameterisation technique that was tested makes use of the generalised Fourier series to decompose the time series into a number of orthogonal functions. This novel feature extraction technique uses the Fourier coefficients for a particular orthogonal basis to represent the time series. The third parameterisation technique that was tested was to try and capture information about the dynamic process whose output is the time series $f_i(t), i = 1, \ldots, n$. This so-called global dynamical model approach parameterises the time series in terms of the coefficients of a set of differential

equations. The fourth and final parameterisation technique that was tested was the EMG histogram as introduced by Zardoshti-Kermani et al. [1995] (see Section 1.4.1). Each of these different parameterisation approaches will now be examined in more detail.

4.1 Envelope Maxima

The simplest parameterisation used during this research was to extract the maximum value of the time series envelope. This is accomplished by firstly removing any line noise or DC offset that may be present in the EMG signal. The resulting clean signal is then rectified. A tenth order polynomial is then fitted to the signal peaks, with a peak being defined as a signal value whose two neighbours are of lesser or equal value. The resulting maxima of the polynomial envelope is then used to characterise the signal.

The fitting of the polynomial is necessitated by the stochastic nature of the EMG signal (see Figure 3.1 through to Figure 3.6). If such a smoothing process were not used then it would be possible for a signal spike entirely unrepresentative of the magnitude of the EMG signal as a whole to be chosen as the characteristic feature. The least squares fitting of the polynomial ensures that the envelope maxima feature is a fair representation of the magnitude of the time series as a whole. This is due to the fact that with ten corner points to manipulate, the least squares process is able to construct a signal envelope that follows the general amplitude trends of the EMG signal without being unduly skewed by aberrant spikes or other such anomalies. A sample set of rectified data and its associated fitted polynomial are shown in Figure 4.1. Such a parameterisation results in each grasp being characterised by an n-tuple of real numbers $(y_1, y_2, \ldots, y_n)$.

4.2 Orthogonal Decomposition

The concept of a vector space containing the set of all real or complex n-tuples can be expanded to include spaces containing sets of functions. There are a number of classes of function that are of great practical value when considered as a space. By abstracting various geometrical concepts into such spaces it is possible to uncover useful facts about both the space and its member functions. One example of such a space is $C[a, b]$, which is the set of all continuous real-valued functions on the interval $[a, b]$.

The time series $f_i(t), i = 1, \ldots, n$, all belong to a space closely related to

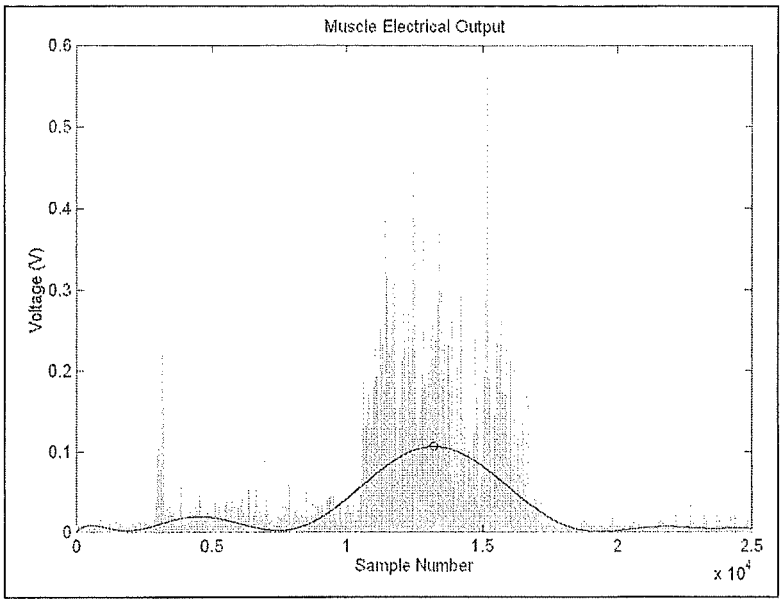

Figure 4.1: Location of the EMG Signal Envelope Maxima. The Circle Indicates the Position of the Maxima.

$C[a,b]$ known as $L_2[a,b]$, or more specifically $L_2[0,T]$.[1]

Examination of Figure 3.1 through to Figure 3.6 or consideration of the physical processes responsible for the generation of the EMG signal should be sufficient to prove that the EMG signal has a finite energy content. This finite energy content ensures the existence of the integral:

$$\int_0^T f^2(t)\,dt$$

which is sufficient to grant the time series $f_i(t), i = 1,\ldots,n$, membership to $L_2[0,T]$.

It is elementary in functional analysis to show that L_2 contains an orthonormal basis $\{g_1(t), g_2(t), \ldots\}$ [Griffel 1981, p. 186]. This fact allows the time series $f_i(t), i = 1,\ldots,n$ to be represented as a linear combination of these basis functions:

$$\begin{aligned} f_i(t) &= \sum_{j=1}^{\infty} a_{ij} g_j(t) \\ &= a_{i1} g_1(t) + a_{i2} g_2(t) + \cdots \end{aligned} \tag{4.1}$$

[1]Readers unfamiliar with the origin of L_2 are advised to read Appendix B.

This series is called the generalised Fourier series of $f_i(t)$ and its coefficients $a_{i1}, a_{i2}, \ldots$ are called the Fourier constants of $f_i(t)$ with respect to the orthogonal basis $g_j(t)$. The Fourier constants are easily found by multiplying both sides of Equation (4.1) by $g_m(t)$ (where m is a fixed number), integrating over $0 \leq t \leq T$, and noting that the integral of the product $g_i(t)g_m(t)$ may be written as $\langle g_i, g_m \rangle$. This gives:

$$\begin{aligned} \langle f_i, g_m \rangle &= \int_0^T f_i(t) g_m(t) dt \\ &= \int_0^T \left(\sum_{j=1}^{\infty} a_{ij} g_j(t) \right) g_m(t) dt \\ &= \sum_{j=1}^{\infty} a_{ij} \langle g_j(t), g_m(t) \rangle \end{aligned} \tag{4.2}$$

The orthogonality of the basis $g_j(t)$ ensures that all of the integrals in Equation (4.2) are zero except when $m = j$, thus:

$$a_{im} = \frac{\langle f_i, g_m \rangle}{\langle g_m, g_m \rangle}$$

It is of course impractical to include an infinite number of terms in Equation (4.1). Only enough terms need to be added to bring the approximate representation of the function $f_i(t)$ "sufficiently close" to the exact function. Just what constitutes sufficiently close depends on the problem to be solved.

In the case of the parameterisation of the EMG signal, an accurate representation using the generalised Fourier series is neither required nor wanted. The objective of parameterising the time series is to enable discrimination between EMG signals from different grasp classes on the basis of a few succinct representative values. Therefore only a few terms need be included in the generalised Fourier representation. It is through the proper selection of the orthogonal basis functions $g_j(t)$ that high rates of discrimination between the grasp classes may be achieved.

Two orthogonal bases were tested during the course of this research: the Legendre polynomials, and the wavelet packets associated with the scaling functions of the Haar wavelet.

4.2.1 Legendre Polynomials

The Legendre polynomials $\{P_n\}$ are a classical set of orthogonal functions that are a result of solving Legendre's differential equation:

$$(1 - x^2)y'' - 2xy' + n(n+1)y = 0 \tag{4.3}$$

where n is a given real number. Legendre's differential equation arises in a wide variety of physical problems, particularly in boundary value problems for spheres. The Legendre polynomials may be found by using the Gram-Schmidt process to orthogonalise the powers $\{1, x, x^2, \ldots\}$ in the space $C[-1, 1]$. It is easy to show that:

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= \frac{1}{2}(3x^2 - 1) \end{aligned}$$

Finding the successive powers of x using the Gram-Schmidt process becomes a little harder. A simpler method is to solve Equation (4.3) using a power series expansion[2]. The result is the so-called recursion formula:

$$P_n(x) = \sum_{m=0}^{M} (-1)^m \frac{(2n-2m)!}{2^n m!(n-m)!(n-2m)!} x^{n-2m}$$

where $n = 1, 2, \ldots$ and $M = n/2$ or $M = (n-1)/2$, whichever is an integer. This recursion formula allows the generation of any Legendre polynomial based on those that have come before it. Figure (4.2.1) illustrates the first five Legendre polynomials.

4.2.2 The Haar Wavelet

The Haar wavelet is one of an extensive class of functions known collectively as wavelets. Wavelets have found great favour in the last few years in a variety of fields; from astrophysics and fluid dynamics to signal and image processing. The great advantage of wavelets over other signal processing techniques is in their ability to extract time-frequency information from a signal.

Consider the following musical analogy [Kaiser 1994]. Suppose that a frequency analysis of a piece of music is required. The piece as perceived by the eardrum may be accurately modelled by a function $f(t)$, representing the air pressure on the eardrum as a function of time. If the music consists of a single steady note with a fundamental frequency ω_1, the $f(t)$ is periodic and the best description of its frequency content is the Fourier series. If the music is a series of such notes or a melody, then it is not periodic in general and the Fourier series cannot be used directly. The Fourier transform $\hat{\omega}(t)$ of $f(t)$ may be computed instead. However, the use of the Fourier transform is flawed

[2]See [Kreyszig 1988, p. 193] for a well written explanation of this process.

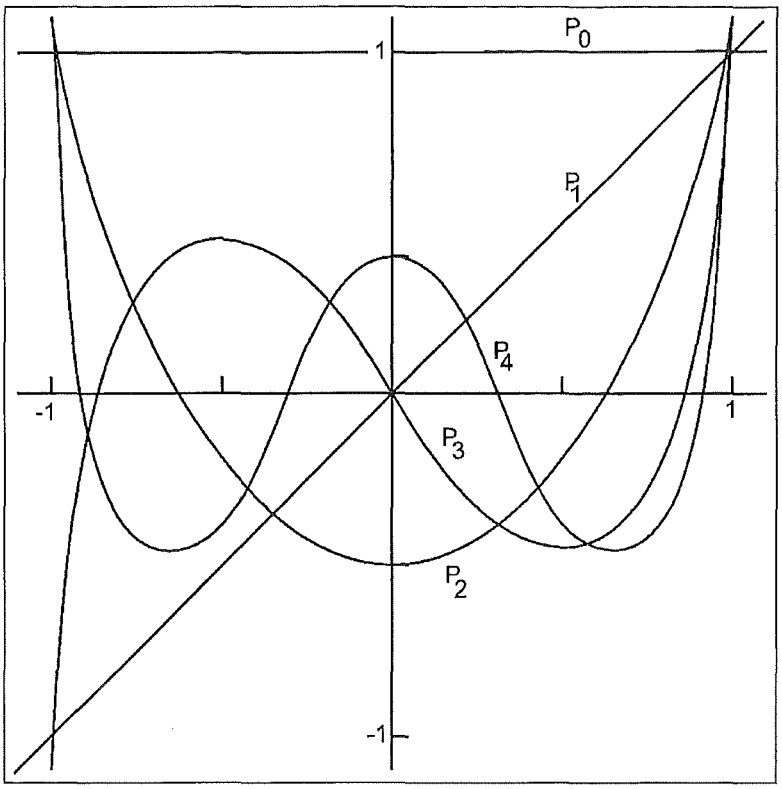

Figure 4.2: The First Five Legendre Polynomials.

from a practical point of view. To compute $\hat{f}(\omega)$ we must integrate $f(t)$ over all time and hence $\hat{f}(\omega)$ contains the total amplitude for the frequency ω in the entire piece rather than the distribution of harmonics in each individual note. Thus, if the piece went on for some time, it would be necessary to wait until it was over before computing $\hat{f}(\omega)$, and then the result would be completely uninformative from a musical point of view. The problem is that this approach is incapable of providing the time-frequency information that is so essential to the hearing of music, and is equally important in other applications.

The ability of wavelets to extract time-frequency information makes them ideal as a means not only to analyse images and signals, but also to compress them for transmission. This is done by replacing the signal that is to be compressed at a given scale by the best approximation that can be drawn at that scale using a particular wavelet. By increasing the resolution and moving to finer and finer scales, an increasingly accurate representation of the signal is obtained. Given an agreed protocol between transmitter and receiver consisting of the wavelets to be used and a crude initial signal, all

that needs to be transmitted is the information regarding the changes in the signal from one scale to another. These changes can then be applied to the initial signal and the original signal reconstructed.

The motivation for the creation of the mathematics of wavelets was a problem posed by Paul Du Bois-Reymond in 1873. He constructed a continuous, 2π periodic function of the real variable x whose Fourier series diverged at a given point. This of course flew in the face of Fourier's assertion that any 2π periodic function could be written as the sum of its Fourier series. One of the avenues pursued by mathematicians to solve this conundrum was to try to find an orthogonal system for which the phenomenon discovered by Du Bois-Reymond in the case of the trigonometric system could not happen. It was this approach which was taken by Haar, whose challenge was to find an orthonormal system of functions $\{h_0(x), h_1(x), \ldots, h_n(x)\}$ defined on $[0, 1]$ such that for any function $f(x)$ continuous on $[0, 1]$ the series:

$$\langle f, h_0\rangle h_0(x) + \langle f, h_1\rangle h_1(x) + \cdots + \langle f, h_n\rangle h_n(x) \tag{4.4}$$

converges to $f(x)$ uniformly on $[0, 1]$. Haar's solution to the problem begins with the function $h(x)$, where:

$$h(x) = \begin{cases} 1, & x \in [0, 1/2) \\ -1, & x \in [1/2, 1) \\ 0, & \text{otherwise} \end{cases}$$

For $n \geq 1$, we have $n = 2^j + k$, $j \geq 0$, and $0 \leq k < 2^j$, giving:

$$h_n(x) = 2^{j/2} h(2^j x - k)$$

To complete the set of functions, we define $h_0(x) = 1$ on $[0, 1)$. This makes the series $\{h_0(x), h_1(x), \ldots, h_n(x), \ldots\}$ an orthonormal basis for $L_2[0, 1]$. This approximation of $f(x)$ by the sequence in Equation (4.4) is nothing more than the approximation of a continuous function by step functions whose values are the mean values of $f(x)$ on the appropriate dyadic intervals. The first four wavelet packets associated with the scaling functions of the Haar wavelet are shown in Figure (4.2.2).

4.3 Global Dynamical Models

The envelope maxima and orthogonal decomposition approaches to signal parameterisation seek to reduce the dimensionality of the time series $f_i(t), i = 1, \ldots, n$, through the calculation of a small number of representative features. Providing that the features from each grasp class differ sufficiently from one

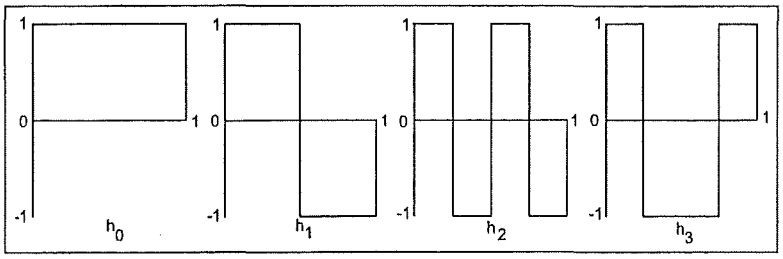

Figure 4.3: The First Four Haar Wavelets.

another then it is possible to discriminate between the grasp classes on the basis of these features.

The global dynamical model approach to signal parameterisation differs from the envelope maxima and orthogonal decomposition techniques in that it seeks to parameterise the dynamical process that produced the time series rather than parameterise the time series themselves [Kadtke 1995]. This technique has its origins in nonlinear dynamical systems theory, and is able to capture nonlinear signal information (correlations). The time series $f_i(t), i = 1, \ldots, n$, are represented using a series of time-delay state space vectors:

$$f_i(t), f_i(t-\tau), f_i(t-2\tau), \ldots, f_i(t-(D-1)\tau)$$

for some embedding dimension D and some time delay τ. This state space representation is essentially nothing more than the discrete time representation of the original time series $f_i(t)$, and is identical to the original sampled signal for sufficient D. To this data is fitted a set of model differential equations:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}(\mathbf{x}, t)$$

The key assumption in this representation is that if the data generator was a continuous dynamical system, then a dynamical model of the correct class will be able to capture all of the essential dynamical information of the system. In practice, a model is constructed by assuming a general expansion for the D functions $\mathbf{F}(\mathbf{x}, t)$ using some set of basis functions $\phi(\mathbf{x}(t))$. These basis functions are not necessarily orthogonal, but they are usually chosen as such for the sake of simplicity.

For the case where $\mathbf{F}(\mathbf{x}, t) = \mathbf{F}(\mathbf{x})$ (that is time independent) all polynomials up to some order P may be used so that:

$$\phi_j(\mathrm{x}(t)) = x(t)^k x^m(t-\tau) \ldots x^q(t-(D-1)\tau)$$

where $0 \leq k, m, q, \ldots \leq P$ and $k + m + q + \ldots \leq P$. This allows us to write for each of the $1 \leq m \leq D$ equations:

$$\begin{aligned} F_m(x,t) &= a_{0,m} + a_{1,m}x(t) + a_{2,m}x(t-\tau) + a_{3,m}x^2(t) + \\ &\quad a_{4,m}x^2(t-\tau) + a_{5,m}x(t)x(t-\tau) + \ldots \\ &= \sum a_{j,m}\phi(\mathbf{x}(t)) \end{aligned}$$

with $j = 1, \ldots, N$. The expansion includes all $N = (P+D)!/P!D!$ cross terms. The unknown model coefficients $a_{j,m}$ are determined by constructing a general linear least squares problem for each F_m. A discrete time approximation to such a representation may be written as:

$$y_k = a_1 y_{k-1} + a_2 y_{k-2} + a_3 y_{k-1}^2 + a_4 y_{k-1}y_{k-2} + a_5 y_{k-2}^2 + a_6 \upsilon_k + \epsilon \qquad (4.5)$$

where y_k is the time series value at sample k, y_{k-1} and y_{k-2} are the time series values at samples $k-1$ and $k-2$ respectively, υ_k is the value of the system input signal at time sample k, and ϵ is the residual error caused by the finite number of terms in the approximation. The number of terms in Equation (4.5) is kept low because most of the signal power is distributed in the low order coefficients, and because the inclusion of a large number of terms would defeat the purpose of parameterising the signal.

Taking $\upsilon_k = 1$ (the unit step) Equation (4.5) can be rewritten as:

$$\begin{aligned} y_k &= \mathbf{\Phi}_k^T \mathbf{a} \\ \mathbf{\Phi}_k^T &= [y_{k-1}\ y_{k-2}\ y_{k-1}^2\ y_{k-1}y_{k-2}\ y_{k-2}^2\ 1] \\ \mathbf{a}^T &= [a_1\ a_2\ a_3\ a_4\ a_5\ a_6] \end{aligned}$$

A recursive least squares algorithm can be formulated for the determination of the coefficients $\mathbf{a}$ such that:

$$\begin{aligned} \mathbf{P}_{k+1} &= \mathbf{P}_k - [1 + \mathbf{\Phi}_{k+1}^T \mathbf{P}_k \mathbf{\Phi}_{k+1}]^{-1} [\mathbf{P}_k \mathbf{\Phi}_{k+1} \mathbf{\Phi}_{k+1}^T \mathbf{P}_k] \\ \mathbf{a}_{k+1} &= \mathbf{a}_k + \mathbf{P}_{k+1}\mathbf{\Phi}_{k+1}[\mathbf{y}_{k+1} - \mathbf{\Phi}_{k+1}^T \mathbf{a}_k] \end{aligned}$$

and where $\mathbf{a}_0 = \mathbf{0}$, and say $\mathbf{P}_0 = c\mathbf{I}$, with $c = 100$. The determination of the coefficients $\mathbf{a}$ using the recursive least squares algorithm [Ljung and Soderstrom 1983] allows the convergence of the coefficients (or otherwise) to be tested. This would not be possible using a nonrecursive solution to the problem.

4.4 EMG Histogram

The EMG histogram method of feature extraction used in this research is a modification of that discussed by Zardoshti-Kermani et al. [1995]. In order

to parameterise the time series $f_i(t), i = 1, \ldots, n$, the axes of the dependant variable (in this case voltage) is split into a number of zones, or bins. A frequency count is then kept for each bin. These counters are used to record the number of times that the value of the time series falls into the range of each bin, as measured at the discrete-time sampling points.

The EMG histogram feature is thus capturing amplitude information about the time series $f_i(t), i = 1, \ldots, n$ in a more thorough fashion than the envelope maxima feature.

4.5 Multivariate Clustering and Classification

Having detailed the methods by which the time series $f_i(t), i = 1, \ldots, n$, may be parameterised, the issue of determining which of these sets of features provides the best discrimination information will now be addressed.

The *hit rate* for a particular feature set is defined as a percentage measure of its ability to allocate a set of features extracted from the time series of a grasp of unknown identity to the correct grasp class. The hit rate is calculated by firstly recording the time series $f_i(t), i = 1, \ldots, n$, for each of the grasps between which discrimination is required. In the case of this research, p sets of n time series were recorded as volunteers grasped each of m objects. From each of the mp sets of n time series, ln parameters were extracted (l from each of the n time series). These parameters are then split into training and test sets such that half of the p sets of ln parameters from each of the m classes were in the training set, and half were in the test set. This separation into training and test sets was done on a random basis. The features from the training set are then used to calculate certain statistical parameters of the grasp classes. These statistical parameters are then used to classify the features contained in the test set. By recording the number of times that each one of the $p/2$ sets of ln parameters from the test set is allocated to the correct grasp class, a percentage measure of success is able to then be calculated.

The l parameters extracted from each of the n time series can be assembled into a $ln \times 1$ column vector. The $ln \times 1$ column vector may be thought of as representing a point in ln-space. This geometrical analogy may be extended to all $p/2$ sets of ln parameters from each of the m grasp classes within the training set, which may then be thought of as forming an ln-dimensional cluster in ln-space.

The question that now arises is how to allocate each of the $p/2$ $ln \times 1$ column vectors of the test set to one of the m ln-dimensional clusters of training data in ln-space. The solution to this problem lies in the selection of

an appropriate *norm*. [3] The simplest norm that may be used is the Euclidean distance $d_e(i), i = 1, \ldots, m$ between the mean of each of the m clusters and the point whose grasp class is unknown.

The use of the Euclidean norm is better understood when considered in the light of the classical two class discrimination problem. Suppose that there are two clusters C_1 and C_2 from which there are a set of reference observations $\mathbf{Y}_1$ and $\mathbf{Y}_2$ which are known to come from C_1 and C_2 respectively. Each of these sets of reference observations is made up of $p/2$ columns of $ln \times 1$ column vectors from each cluster. Using the observations from the reference sets $\mathbf{Y}_1$ and $\mathbf{Y}_2$ it is a simple matter to form the sample mean $ln \times 1$ column vectors $\overline{\mathbf{y}}_1$ and $\overline{\mathbf{y}}_2$. The Euclidean distance between an observation $\mathbf{y}$ whose grasp class is unknown and the two clusters C_1 and C_2 is then given by:

$$d_e(i) = \left\{(\mathbf{y} - \overline{\mathbf{y}})^T(\mathbf{y} - \overline{\mathbf{y}})\right\}^{1/2}$$

for $i = 1, 2$. The observation $\mathbf{y}$ would then be assigned to the cluster whose Euclidean distance $d_e(i), i = 1, 2$ is smallest. The two class discrimination problem may easily be extended to the general m class problem by forming the mean column vectors $\overline{\mathbf{y}}_i, i = 1, \ldots, m$ from the sets of reference observations from the clusters $C_i, i = 1, \ldots, m$. As before an observation whose grasp class is unknown would then be assigned to the cluster whose Euclidean distance $d_e(i), i = 1, \ldots, m$ is smallest.

The simplicity of the Euclidean norm makes it ideal for incorporation into a real-time discrimination algorithm. This is due to the fact that it is a quantity that would be very quickly calculated within the vector math environment of a digital signal processer (DSP) chip. It is however this very simplicity which renders the Euclidean norm unreliable in all but the most simplistic discrimination problems. To illustrate its deficiencies, consider again the classical two class discrimination problem. This time, let each set of $p/2$ reference observations from clusters C_1 and C_2 contain only two coordinates. This allows the two clusters to be plotted on the Cartesian plane along with the observation $\mathbf{y}$ whose grasp class is unknown. For the purpose of this example, let the observation have come from cluster C_2. As can be seen from Figure 4.4, the observation will be allocated to cluster C_1 even though it is both known and apparent from Figure 4.4 that the observation came from cluster C_2. What the Euclidean norm lacks is the ability to take into account the "spread" of the cluster to which the norm is being calculated. This concept of spread is quantified in the statistic known as the *covariance*.

[3] For an explanation of the concept of the norm, see Appendix B.

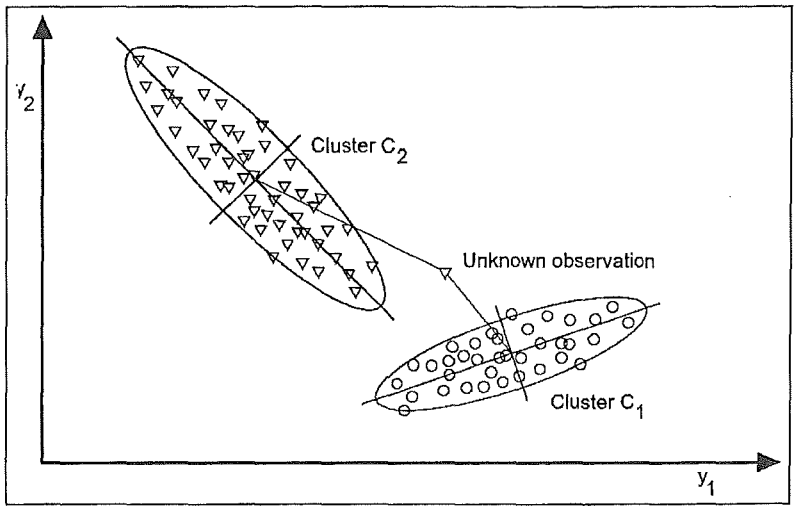

Figure 4.4: An Illustration of the Deficiencies of the Euclidean Norm.

The *variance* σ_y^2 of a set of univariate observations $y_1, y_2, \ldots, y_p$ is defined as:

$$\sigma_y^2 = \frac{\sum_{i=1}^{p}(y_i - \overline{y})^2}{p} \tag{4.6}$$

where $\overline{y}$ is the mean of the y_p values. Variance is simply the average distance of the data points $y_1, y_2, \ldots, y_p$ from the mean. These distances are squared to make them all positive. This concept of variance may easily be extended to include multivariate data. Consider the case of the four clusters of bivariate data shown in Figure 4.5. It may be seen that the elliptical cluster C_1 is aligned so that its major axis lies at approximately 45° to the y_1 variable axis. Thus, given the fact that the y_1 values associated with any particular value of y_2 exhibit less dispersion than the collection of y_2 values in cluster C_1 as a whole, one can make a more precise estimate of a likely y_2 value if the value of y_1 is known rather than not known. The variables y_1 and y_2 are then said to be statistically related. Such a statistical relation is characterised by two quantities; its strength, which is a measure of the degree to which the precision of estimate of y_2 is increased by knowing y_1; and its direction, which gives an indication of the orientation of the cluster. In the case of cluster C_1 in Figure 4.5, small y_1 values are coupled with small y_2 values, and large y_1 values are coupled with large y_2 values. The direction of such a relationship is said to be positive. These concepts of the strength and direction of a statistical relationship are quantified in the statistical property known as the covariance. The covariance $C_{y_1y_2}$ of two sets of variables y_{1i} and y_{2i}, with

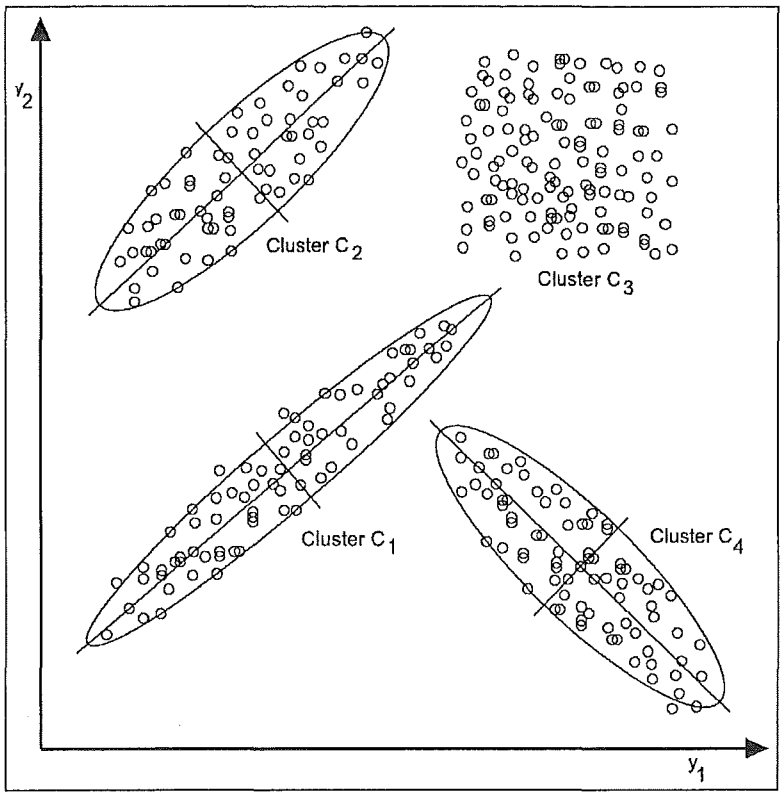

Figure 4.5: An Illustration of the Concept of a Statistical Relation.

$i = 1, \ldots, p/2$ in the case of the sets of training data, is given by:

$$C_{y1y2} = \frac{\sum_{i=1}^{p/2}(y_{1i} - \overline{y_1})(y_{2i} - \overline{y_2})}{p/2}$$

The absolute value of C_{y1y2} increases as the strength of the relation between y_1 and y_2 increases. This point is well illustrated by considering clusters C_1 and C_2 in Figure 4.5. It is clear that a much stronger relation exists between the y_1 and y_2 axis for cluster C_1 than it does for cluster C_2. Assuming that each cluster contains an equal number of observations, the sum $\sum_{i=1}^{p/2}(y_{1i} - \overline{y_1})(y_{2i} - \overline{y_2})$ for cluster C_2 will be greater simply by virtue of the larger average distance between each observation and the cluster mean. This results in cluster C_2 having a larger covariance C_{y1y2} than cluster C_1 whose sum of distances from the mean is smaller. The covariance of clusters C_1 and C_2 will be significantly larger than that of cluster C_3 due to the fact that the axes y_1 and y_2 are essentially unrelated in cluster C_3. Finally, the orientation of the cluster is indicated by the sign of the covariance $C_{y_1y_2}$. Thus clusters C_1

and C_2 will have a positive covariance whilst cluster C_4 will have a negative covariance.

The two statistical properties of variance and covariance are brought together in the so-called covariance matrix S. For an ln-dimensional cluster, the covariance matrix is a square symmetrical $ln \times ln$ matrix containing the variances of the ln axes of the cluster along the diagonal, and with the off-diagonal positions of the matrix being filled with the covariance of the two axes given by the row and column numbers of that particular location. Hence:

$$\mathbf{S} = \begin{bmatrix} \sigma^2_{y_1} & C_{y_1y_2} & C_{y_1y_3} & \cdots & C_{y_1y_{ln}} \\ C_{y_2y_1} & \sigma^2_{y_2} & C_{y_2y_3} & \cdots & C_{y_2y_{ln}} \\ C_{y_3y_1} & C_{y_3y_2} & \sigma^2_{y_3} & \cdots & C_{y_3y_{ln}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{y_{ln}y_1} & C_{y_{ln}y_2} & C_{y_{ln}y_3} & \cdots & \sigma^2_{y_{ln}} \end{bmatrix}$$

The covariance matrix gives a measure of the statistical relation that exists between every possible set of orthogonal axes within a cluster, and also provides a measure of the amount of spread along each axis. It is through the inclusion of the covariance matrix that it is possible to modify the Euclidean norm to compensate for the spread of the cluster to which the norm is being calculated.

The *Mahalanobis* distance d^2_m between an observation $\mathbf{y}$ whose grasp class is unknown and a cluster C_i is given by:

$$d^2_m(i) = (\mathbf{y} - \overline{\mathbf{y}}_i)^T \mathbf{S}_i^{-1} (\mathbf{y} - \overline{\mathbf{y}}_i)$$

where $\overline{\mathbf{y}}_i$ is the $ln \times 1$ mean column vector of the ith cluster, and $\mathbf{S}_i^{-1}$ is the inverse of the covariance matrix of the ith cluster.

By "dividing" the Euclidean norm by the covariance matrix, it is possible to compensate for the statistical uncertainty in the measurement of each point in the cluster. This uncertainty comes about from the fact that the data within a cluster is only a sample from the population, and thus any statistical properties which are calculated from the data within the cluster such as the mean are subject to a degree of uncertainty. By dividing by the covariance matrix, any axes which have a large variance are less heavily weighted in the calculation of the norm than those axes whose variance is smaller.

The use of a norm such as the Mahalanobis distance which is able to take into account the spread of the clusters will improve the hit rate over a simple norm such as the Euclidean distance. However, a far greater improvement in the hit rate is possible by transforming the entire ln-dimensional space in which the clusters reside.

4.5.1 Discriminant Transform

The most popular class of feature space transformations are commonly referred to as *discriminant* transforms. The transformation used during this research was the so-called Fisher-Rao transformation, which is an m cluster adaptation of the two class Fisher discriminant function.

Fisher's discriminant function determines the vector of maximum separation between two clusters of data in ln-space. Suppose that there are two clusters C_1 and C_2 from which there are a set of reference observations $\mathbf{Y}_1$ and $\mathbf{Y}_2$ which are known to come from C_1 and C_2 respectively. Each of these sets of reference observations is made up of $p/2$ columns of $ln \times 1$ column vectors from each cluster. Using the observations from the reference sets $\mathbf{Y}_1$ and $\mathbf{Y}_2$ it is a simple matter to form the sample mean column vectors $\overline{\mathbf{y}}_1$ and $\overline{\mathbf{y}}_2$. Likewise it is a simple matter to form the two covariance matrices $\mathbf{S}_1$ and $\mathbf{S}_2$. Having established these quantities, the objective now is to transform the multivariate $ln \times 1$ observations $\mathbf{y}$ into univariate observations z such that the z values from clusters C_1 and C_2 are separated as much as possible. The motivation for forming a univariate statistic is that a one dimensional vector is far easier to work with than a multivariate statistic or the ln-dimensional observations themselves. Let:

$$\begin{aligned} z &= \sum_{i=1}^{ln} a_i y_i \\ &= \mathbf{a}^T \mathbf{y} \end{aligned}$$

where $\mathbf{y}$ is an $ln \times 1$ observation from clusters C_1 and C_2, $\mathbf{a}$ is a $ln \times 1$ column vector of linear multipliers, and z is the resulting univariate statistic for each point in clusters C_1 and C_2. Fishers discriminant function is that linear combination of the $p/2$ original observations from each cluster which has the largest ratio of variance between the two clusters relative to that of all observations from both clusters considered as a whole. That is, Fishers discriminant function seeks to maximise λ, where:

$$\lambda = \frac{\text{variance between the two clusters}}{\text{variance of the two clusters combined}}$$

The variance between two discrete points is defined in Equation (4.6) to be the distance between them squared. Thus, the variance between the two clusters is in fact the distance between their means squared. The variance of the two clusters combined is simply the variance of the statistic z, as all of the ln-dimensional observations from both clusters are transformed to z values. This enables λ to be rewritten as:

$$\lambda = \frac{(\text{distance between the means of the two clusters})^2}{\text{variance of } z}$$

$$
\begin{aligned}
&= \frac{(\mathbf{a}^T\overline{\mathbf{y}}_1 - \mathbf{a}^T\overline{\mathbf{y}}_2)^2}{\sigma_z} \\
&= \frac{\mathbf{a}^T(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)^T\mathbf{a}}{\mathbf{a}^T\mathbf{S}_p\mathbf{a}}
\end{aligned} \tag{4.7}
$$

The variance of z, $\sigma_z = \mathbf{a}^T\mathbf{S}_p\mathbf{a}$ comes from the fact that z is a linear combination of $\mathbf{a}$ and $\mathbf{y}$. The covariance matrix $\mathbf{S}_p$ is the pooled covariance of clusters C_1 and C_2 and is given by [Neter *et al.* 1988, p. 402]:

$$
\mathbf{S}_p = \frac{(p/2-1)\mathbf{S}_1 + (p/2-1)\mathbf{S}_2}{p/2 + p/2 - 2}
$$

where:

$$
\begin{aligned}
\mathbf{S}_1 &= \frac{1}{p/2-1}\sum_{i=1}^{p/2}(\mathbf{y}_i - \overline{\mathbf{y}}_1)(\mathbf{y}_i - \overline{\mathbf{y}}_1)^T \\
\mathbf{S}_2 &= \frac{1}{p/2-1}\sum_{i=1}^{p/2}(\mathbf{y}_i - \overline{\mathbf{y}}_2)(\mathbf{y}_i - \overline{\mathbf{y}}_2)^T
\end{aligned}
$$

The ratio λ given by Equation (4.7) is maximised by choosing:

$$
\mathbf{a} = c\mathbf{S}_p^{-1}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)
$$

where c is a constant. In the ln-dimensional space of the two clusters C_1 and C_2 such a vector $\mathbf{a}$ defines the direction of maximal cluster separation in the sense that the means of the projections of the two clusters are maximally apart relative to the variance of the projections around their respective means. Choosing $c = 1$ leads to:

$$
\begin{aligned}
z &= \mathbf{a}^T\mathbf{y} \\
&= \{\mathbf{S}_p^{-1}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)\}^T\mathbf{y}
\end{aligned} \tag{4.8}
$$

Equation (4.8) is known as Fisher's discriminant function. Figure (4.6) shows how the Fisher's discriminant function works for the case where $ln = 2$. Suppose now that instead of two clusters, there are m clusters $C_i, i = 1, \ldots, m$. As for the two group case there are a set of reference observations $\mathbf{Y}_i, i = 1, \ldots, m$. Each of these sets of reference observations is made up of $p/2$ columns of $ln \times 1$ column vectors from each cluster. Using these reference observations it is a simple matter to form the $ln \times 1$ mean column vectors $\overline{\mathbf{y}}_i, i = 1, \ldots, m$. What Rao's generalisation of Fishers discriminant function does is determine the multiple axis along which the m clusters are separated as much as possible. This is done in the same manner as the two cluster case

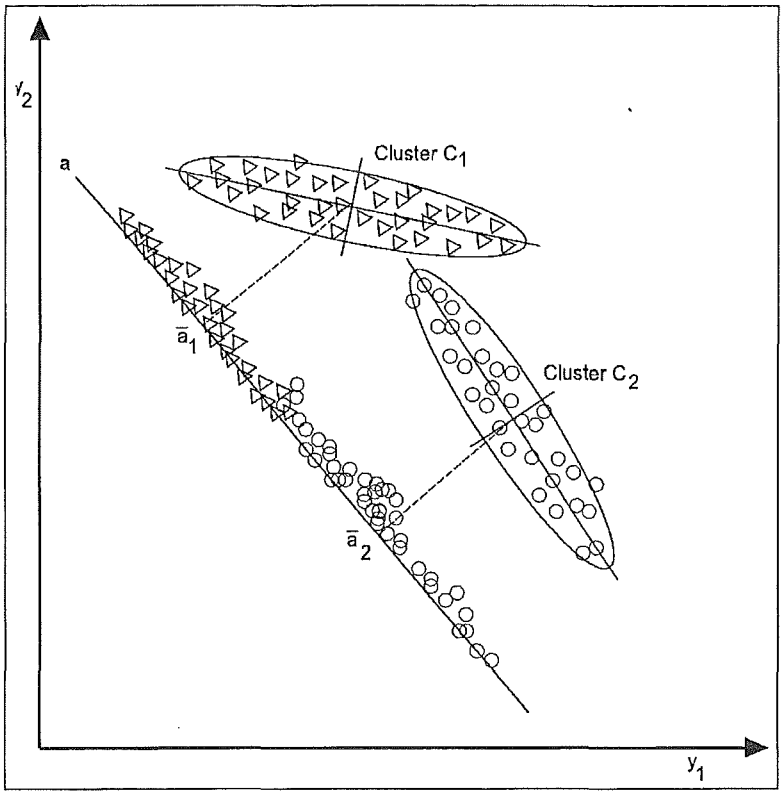

Figure 4.6: An Illustration of the Working of Fisher's Discriminant Function.

by forming the univariate statistic $z = \mathbf{a}^T\mathbf{y}$. Recall that in the two cluster case the objective was to maximise the ratio λ given by:

$$\begin{aligned} \lambda &= \frac{\text{variance between the two clusters}}{\text{variance of both clusters combined}} \\ &= \frac{\mathbf{a}^T(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)^T\mathbf{a}}{\mathbf{a}^T\mathbf{S}_p\mathbf{a}} \end{aligned} \tag{4.9}$$

In the multigroup case the objective is exactly the same. However now the variance between clusters (as measured by the squared distance between the cluster means) will be a matrix due to their being more than two clusters. This between clusters covariance matrix is given the symbol $\mathbf{B}$ and is given by:

$$\mathbf{B} = \frac{1}{m-1}\sum_{i=1}^{m}(\overline{\mathbf{y}}_i - \overline{\mathbf{y}})(\overline{\mathbf{y}}_i - \overline{\mathbf{y}})^T$$

where $\overline{\mathbf{y}}_i$ is the mean column vector of the ith cluster, and $\overline{\mathbf{y}}$ is the mean of all of the $p/2$ data points from each of the m clusters. The between clusters

covariance matrix $\mathbf{B}$ gives a measure of the spread of the individual cluster means compared to the mean of all observations considered as a whole.

In addition to the between groups covariance matrix, a statistic able to measure the variance of all of the clusters combined is required for use in Equation (4.9). In the two cluster case use was made of the pooled covariance matrix $\mathbf{S}_p$. In the m cluster case:

$$\mathbf{S}_p = \frac{\mathbf{W}}{(mp/2) - m}$$

where:

$$\mathbf{W} = \frac{1}{p/2 - m} \sum_{i=1}^{m} (p/2 - 1)\mathbf{S}_i$$

It may be noted that the value of $\mathbf{a}$ that maximises the ratio:

$$\begin{aligned} \lambda &= \frac{\text{variance between the two clusters}}{\text{variance of both clusters combined}} \\ &= \frac{\mathbf{a}^T\mathbf{B}\mathbf{a}}{\mathbf{a}^T\mathbf{S}_p\mathbf{a}} \end{aligned}$$

also maximises the ratio:

$$\frac{\mathbf{a}^T\mathbf{B}\mathbf{a}}{\mathbf{a}^T\mathbf{W}\mathbf{a}}$$

Thus $\mathbf{W}$ can be used in place of $\mathbf{S}_p$. The question that now arises is how to maximise the ratio λ. Given that a maxima is a point of zero gradient, and that gradient is the first derivative, all that is necessary is to differentiate λ with respect to $\mathbf{a}$ and set it equal to zero:

$$\begin{aligned} \frac{\partial\lambda}{\partial\mathbf{a}} &= \frac{\partial}{\partial\mathbf{a}}\left(\frac{\mathbf{a}^T\mathbf{B}\mathbf{a}}{\mathbf{a}^T\mathbf{W}\mathbf{a}}\right) \\ &= \frac{\mathbf{a}^T\mathbf{W}\mathbf{a}(\mathbf{B}^T\mathbf{a} + \mathbf{B}\mathbf{a}) - \mathbf{a}^T\mathbf{B}\mathbf{a}(\mathbf{W}^T\mathbf{a} + \mathbf{W}\mathbf{a})}{(\mathbf{a}^T\mathbf{W}\mathbf{a})^2} \\ &= \frac{\mathbf{a}^T\mathbf{W}\mathbf{a}(2\mathbf{B}\mathbf{a}) - \mathbf{a}^T\mathbf{B}\mathbf{a}(2\mathbf{W}\mathbf{a})}{(\mathbf{a}^T\mathbf{W}\mathbf{a})^2} \\ &= \frac{2\{\mathbf{a}^T\mathbf{W}\mathbf{a}(\mathbf{B}\mathbf{a}) - \mathbf{a}^T\mathbf{B}\mathbf{a}(\mathbf{W}\mathbf{a})\}}{(\mathbf{a}^T\mathbf{W}\mathbf{a})^2} \\ &= \frac{2\{\mathbf{B}\mathbf{a} - \frac{\mathbf{a}^T\mathbf{B}\mathbf{a}}{\mathbf{a}^T\mathbf{W}\mathbf{a}}\mathbf{W}\mathbf{a}\}}{\mathbf{a}^T\mathbf{W}\mathbf{a}} \\ &= \frac{2\{\mathbf{B}\mathbf{a} - \lambda\mathbf{W}\mathbf{a}\}}{\mathbf{a}^T\mathbf{W}\mathbf{a}} \\ &= \mathbf{B}\mathbf{a} - \lambda\mathbf{W}\mathbf{a} \\ &= 0 \end{aligned}$$

This is of course the classic generalised eigenvalue problem. If $\mathbf{a}$ is chosen to maximise the λ ratio, that value of $\mathbf{a}$ turns out to be an eigenvector $\mathbf{a}_1$ corresponding to the largest eigenvalue c_1 of $\mathbf{W}^{-1}\mathbf{B}$. Having done this, a second linear combination of the original variables which has the next largest λ ratio may be sought. If this second linear combination is constrained to be orthogonal to the first, then required solution turns out to be the eigenvector $\mathbf{a}_2$ corresponding to the second largest eigenvalue c_2. This process may be repeated to determine additional linear combinations. All that needs to be done is to perform an eigenanalysis of $\mathbf{W}^{-1}\mathbf{B}$.

Essentially the Fisher-Rao transformation treats the means of the individual clusters $C_i, i = 1, \ldots, m$ as the data points of an ellipsoidal distribution centred on the mean of the $p/2$ points from each of the m clusters, and with a variance of $\mathbf{S}_p$. The transformation produces eigenvectors corresponding to the axes of the hyperellipsoid of greatest variance. The number of eigenvectors produced is determined by the dimensionality of the space rather than the number of clusters within that space. This concept is illustrated in Figure 4.7. It is very important to note that the eigenvectors of $\mathbf{W}^{-1}\mathbf{B}$ may

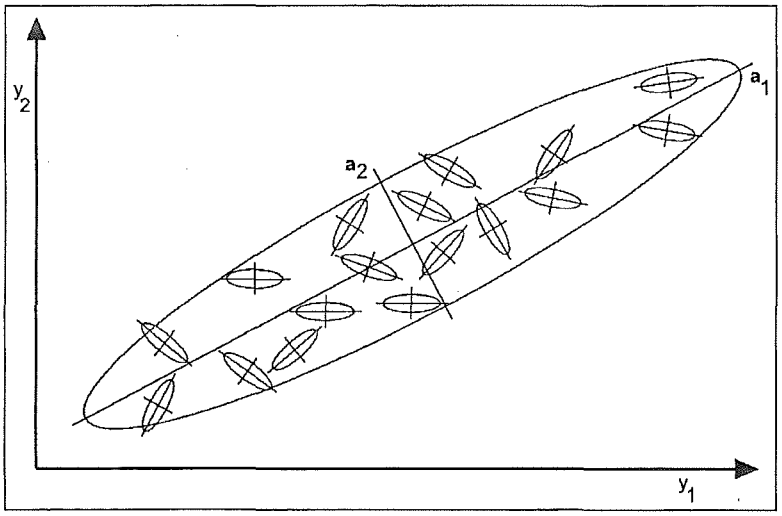

Figure 4.7: An Illustration of the use of the Fisher-Rao Transformation.

not be orthogonal. This is due to the fact that the matrix $\mathbf{W}^{-1}\mathbf{B}$ cannot be proven to be symmetric. However, even in the case where the eigenvectors are not orthogonal, they are still linearly independent.

Having calculated the eigenvector transformation matrix, it is a simple matter of transforming all the $p/2$ points from each of the $C_i, i = 1, \ldots, m$ clusters into the new discriminant space. This is done by applying the trans-

formation:

$$\mathbf{Z} = \mathbf{A}^T\mathbf{Y}$$

where $\mathbf{A}^T$ is a matrix whose rows are the eigenvectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{ln}$ of $\mathbf{W}^{-1}\mathbf{B}$, and $\mathbf{Y}$ is the set of all reference observations $[\mathbf{Y}_1\mathbf{Y}_2 \cdots \mathbf{Y}_m]$ from each cluster. The matrix $\mathbf{Z}$ may of course be partitioned according to the partitioning of $\mathbf{Y}$ so as to provide the original group identities for the representations in the discriminant space.

The advantage of this new transformed space is that the coordinate axes are those of maximal separation between the clusters. The Euclidean norm can then be used in this new transformed space with much greater success that in the original, untransformed feature space. In fact, the Euclidean norm within the transformed space is the same as the Mahalanobis norm. It is the fact that the transformation eigenvectors are not orthogonal that contributes to an improved hit rate.

Chapter 5

Experimental Results

Having discussed the physiological origins of the EMG signal, the manner in which it is recorded, the mathematical techniques used to parameterise it, and the statistical processes used to classify the resulting features, it is now time to examine the classification results that were obtained for each of the parameterisation techniques that were tested.

5.1 Envelope Maxima

The baseline measure of feature performance is that of the envelope maxima feature. The minimum, mean, and maximum hit rates obtained from the data gathered during the clinical trial using the Euclidean, Mahalanobis, and the transformed Euclidean norms (recall that the transformed Euclidean norm is simply the Euclidean norm used in the Fisher-Rao transformed feature space) are listed in Table 5.1. Note that the Mahalanobis norm performs better than

	Hit Rate (%)		
Norm	Minimum	Mean	Maximum
Euclidean	55.3	66.7	78.4
Mahalanobis	70.3	74.9	86.8
Transformed Euclidean	65.1	73.7	85.3

Table 5.1: Envelope Maxima Hit Rates

the Euclidean, due to the fact that it is able to take into account the variance of the grasp class clusters. What is unexpected is that the Mahalanobis norm also performed better than the more sophisticated transformed Euclidean norm. This is most likely due to the fact that there are simply not enough

degrees of freedom within the grasp class clusters for the eigenvectors of the Fisher-Rao transformation to be much different from the original orthogonal coordinate system.

As a baseline measure of performance, the envelope maxima hit rates are comparatively stable across the ten volunteers as evidenced by the spread of hit rates. The hit rates are however less than those reported by Vuskovic [1995] in which he reported hit rates of 94 and 90% for the two subjects involved. This discrepancy between the two sets of results can be attributed to Vuskovic's superior electrode placement scheme. His real-time signal display allowed for the electrodes to be moved about until the best position was found. In this research however, the electrode position was not optimised in such a way. In addition, it is quite possible that Vuskovic's small sample size may have worked in his favour. Examination of the results obtained over the course of this research show significant differences in performance of the various signal parameterisation techniques from one volunteer to the next. Volunteers whose data performed well under one parameterisation technique invariably performed well under them all. If Vuskovic's two volunteers were people whose EMG data was superior to that of the population as a whole, this would help to account for the discrepancies in hit rates.

5.2 Orthogonal Decomposition

The orthogonal decomposition parameterisation technique rewrites each channel of EMG data gathered as the sum of a series of orthogonal functions, with coefficients representing the "amount" of each of orthogonal function contained within the EMG signal. The two sets of orthogonal functions used during the course of this research were the Legendre polynomials, and the wavelet packets associated with the scaling functions of the Haar wavelet (the Haar wavelet for short). The first 50 functions from each of these sets of orthogonal functions were used, resulting in each channel of EMG data being parameterised by 50 coefficients.

5.2.1 Linear Classification

Recall from Section 3.6 that a single complete data run consists of features extracted from 100 records of four channels of EMG data recorded from each of the 6 grasp classes between which discrimination is desired. This data is split in half on a random basis such that 50 sets of features from each grasp class are contained in the training data set, and 50 sets of features from each grasp class are contained in the test data set. The data from the

training data set is used to calculate certain statistical features about each grasp class. These statistics are then used in classifying the data in the test set. This process of randomly splitting the data into training and test sets is performed 500 times, resulting in a stable estimate of the hit rate. The hit rates calculated represent the mean hit rate obtained across all grasp classes using the Euclidean, Mahalanobis, and transformed Euclidean norms (the Euclidean norm used in the Fisher-Rao transformed feature space).

If all 50 features from each of the four channels was to be included in the classification procedure, it would take several weeks to complete a single run. In order to reduce the computation time it was decided to only use the first ten features extracted from each channel. This results in each grasp being characterised by 40 features rather than 200.

In order to examine the effect that increasing the dimensionality of the feature space would have upon the hit rate, it was decided to calculate hit rates as each of the first ten features were added sequentially. During classification, the hit rate was first calculated for the case where only the first of the ten features from each channel was included, resulting in a four dimensional feature space. The hit rate was then recalculated with the first and second features included, resulting in an eight dimensional feature space. The addition of features and the recalculation of the hit rate continued until the first ten features had been added, resulting in a 40 dimensional feature space. This process of adding features in their "natural" order is referred to as linear classification.

5.2.2 Ranked Classification

In addition to examining the effect that the number of features would have on the hit rate, the effect of the order of the ten features was also investigated. In order to do this, it was decided to try an alternative feature ranking. This alternative ranking orders the first 50 features from best to worst according to the sum of the squared Euclidean distances between the mean of each grasp class and that of every other grasp class. Thus, a feature whose sum of squared distances is large is ranked a better discriminator than a feature whose sum of squared distances is low. The logic behind this measure is that it is desirable to have the means of the grasp classes spread out as far as possible through the feature space, as this minimises the chances of two or more grasp classes overlapping. Once the first 50 coefficients had been ranked using this technique, the first ten of these were then classified in a manner identical to that used for the first ten linearly ranked features. That is, the hit rate was calculated for the best feature, then the first and second best features, and so on until hit rates had been calculated as each of the top

ten features were added.

5.2.3 Legendre Polynomials

The effect that increasing the dimensionality of the feature space has is shown in Figures 5.1 through to Figure 5.3, which show the variation in hit rate as measured using the Euclidean, Mahalanobis, and transformed Euclidean norms.

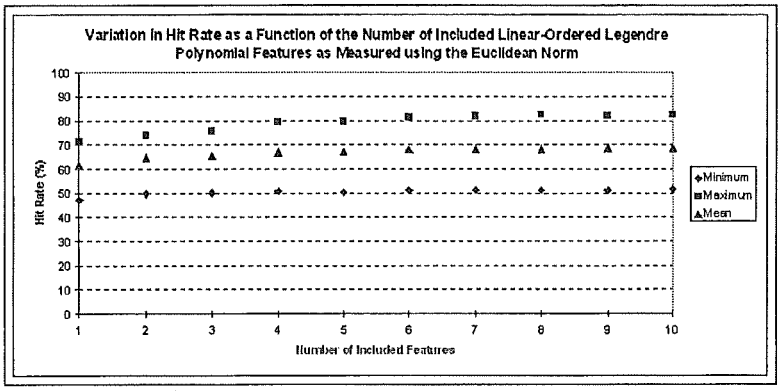

Figure 5.1: Variation in Hit Rate with the Number of Included Linear-Ordered Legendre Polynomial Features as Measured using the Euclidean Norm.

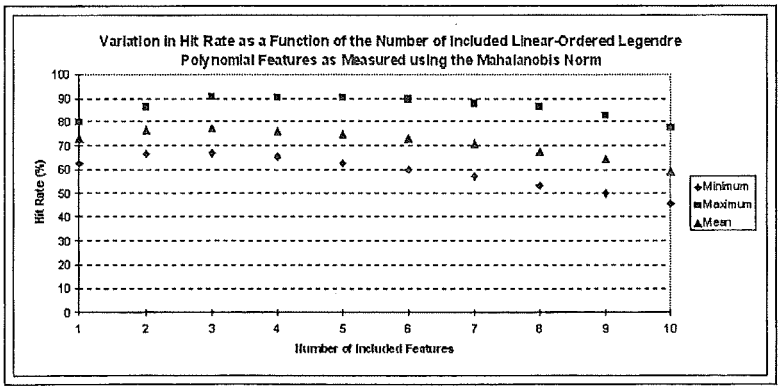

Figure 5.2: Variation in Hit Rate with the Number of Included Linear-Ordered Legendre Polynomial Features as Measured using the Mahalanobis Norm.

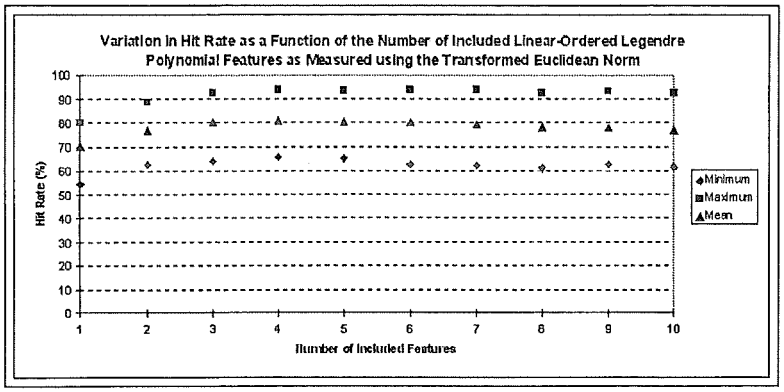

Figure 5.3: Variation in Hit Rate with the Number of Included Linear-Ordered Legendre Polynomial Features as Measured using the Transformed Euclidean Norm.

As can be seen from Figure 5.1 through to Figure 5.3, increasing the dimensionality of the feature space has a marked effect on the hit rate—regardless of which norm is being considered.

The inclusion of only the first Legendre polynomial feature (P_0) results in mean hit rates of 61.6%, 73.0%, and 70.3% using the Euclidean, Mahalanobis, and transformed Euclidean norms respectively. If these hit rates are considered as reference measures of performance, then it is apparent that the addition of the second Legendre polynomial feature results in an improvement in the reference hit rates. Providing that the second Legendre polynomial feature differs significantly between grasp classes and is relatively stationary within each class, then its inclusion in the feature space brings additional information that may be used to help discriminate between grasp classes. Similarly the addition of higher order Legendre polynomial features will improve the hit rates further, providing that these features bring additional positive information to the discrimination problem. This trend may be seen in Figures 5.1 through to 5.3, where the incorporation of the third Legendre polynomial feature results in improvements to the hit rates.

It is important to note however that each of these figures has a point on it after which the inclusion of additional features has the effect of reducing the hit rate. This reduction in hit rate continues as more and more features are added until the hit rates reach a point where they actually fall below their reference levels. This effect may be attributed to the fact that the incorporation of these additional features is not bringing any new information to the discrimination problem. In order to understand how this is possible, consider the case of the one dimensional data shown in Figure 5.4. As can

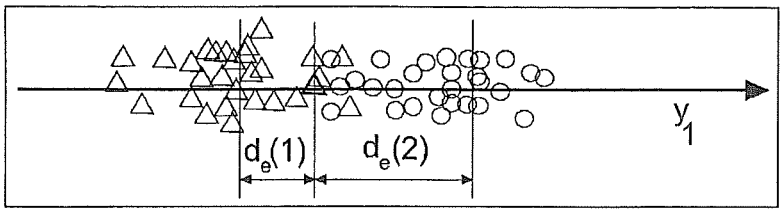

Figure 5.4: The Two-Class Discrimination Problem with Univariate Data. Note that the data points from the two classes have been spread out across the y axis for clarity only, and this distribution should not be interpreted as a second data dimension.

be seen, the features from each of the two classes overlap a small amount, but the unknown observation is still able to be correctly classified on the basis of the Euclidean norms $d_e(1)$ and $d_e(2)$. If however a second set of features is added that provide no additional discrimination information (such as uniform white noise), the probability of the unknown observation being incorrectly classified becomes much greater. This is illustrated in Figure 5.5 in which the random value of the y_2 feature results in the possibility of Euclidean norms between the two classes becoming identical, thus leading to the misclassification of the unknown observation. Note that each of the

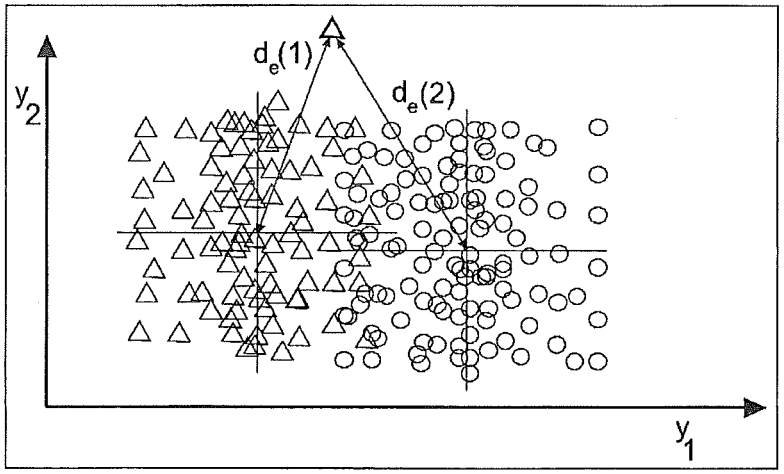

Figure 5.5: The Two-Class Discrimination Problem with Noisy Bivariate Data.

three norms are susceptible to this effect to a different degree. For instance, the transformed Euclidean norm is effected very little by the addition of a large number of features. Its hit rate reaches a peak with the inclusion of

four features, and then decreases by approximately 2-3% as the number of features is further increased to ten. The Mahalanobis norm however, is far more susceptible to this phenomenon. The maximum hit rate as measured using the Mahalanobis norm peaks with the inclusion of three features, but then falls by approximately 12 percentage points as all ten features are added. The norm with the best hit rate performance on the whole is the transformed Euclidean. It provides the best maximum hit rate of all norms of 94%, as well as the best mean hit rate of 80%.

The effect that changing the order in which features are added to the feature space is shown in Figure 5.6 through to Figure 5.8. These show the variation in hit rate as measured using the Euclidean, Mahalanobis, and transformed Euclidean norms as the best ten features are added in ranked order.

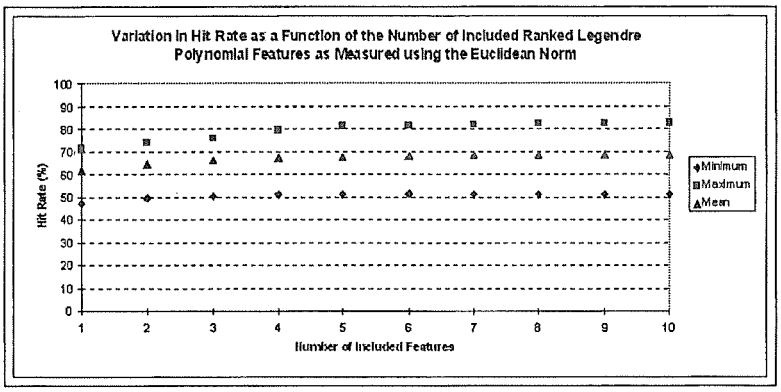

Figure 5.6: Variation in Hit Rate with the Number of Included Ranked Legendre Polynomial Features as Measured using the Euclidean Norm.

The order in which the Legendre polynomial features are added to the feature space has very little effect on the resulting hit rates. This is due to the fact that most of the signal power (and corresponding discriminatory ability) is contained in the low order features corresponding to the low order Legendre polynomials. This is well illustrated by Table 5.2. As can be seen from Table 5.2, the procedure used to rank the first 50 Legendre polynomial features selected the first ten features almost without exception. Furthermore, the ranking procedure did not alter the ordering of these first ten features very much at all. This provides conclusive proof that, in the case of the Legendre polynomial features, the ability to capture signal power information, and thus discriminatory ability, is concentrated in the lower order features. This also implies that feature ordering will only really produce significant

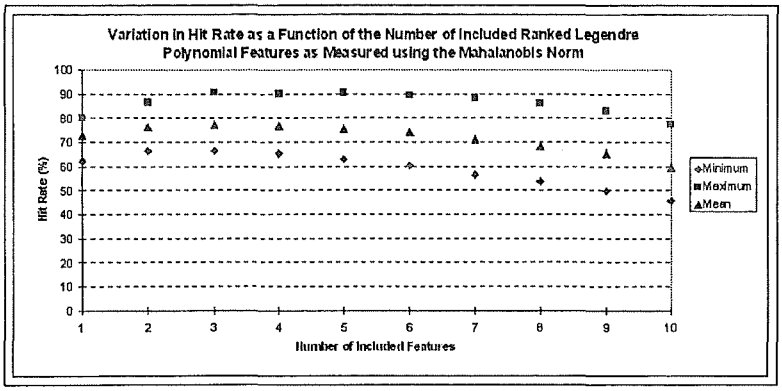

Figure 5.7: Variation in Hit Rate with the Number of Included Ranked Legendre Polynomial Features as Measured using the Mahalanobis Norm.

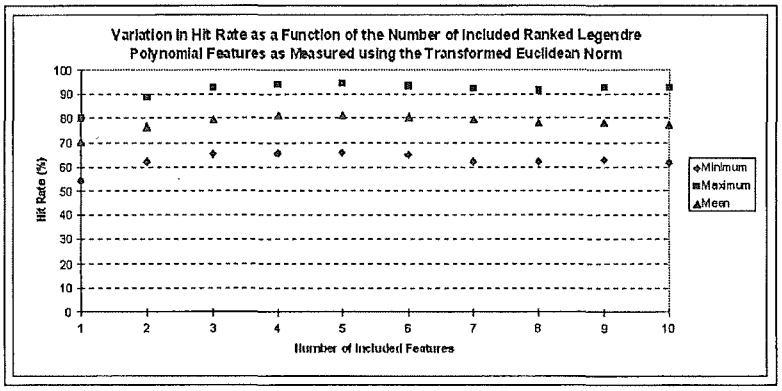

Figure 5.8: Variation in Hit Rate with the Number of Included Ranked Legendre Polynomial Features as Measured using the Transformed Euclidean Norm.

improvements in hit rates for those sets of orthogonal functions whose ability to capture signal power information is concentrated in the higher order functions.

The hit rates obtained using the Legendre polynomial features are summarised in Table 5.3.

		Feature Ranking									
		1	2	3	4	5	6	7	8	9	10
Feature	Other						1		1	1	5
	10								1	4	1
	9							2	3	4	1
	8						1	2	1		3
	7				1	1	5	1	2		
	6				1	6		2			
	5				1	1	2	3	2	1	
	4			7	3						
	3		1	2	4	2	1				
	2		9	1							
	1	10									

Table 5.2: Legendre Polynomial Feature Ranking Frequencies.

Norm	Order	Hit Rate (%)		
		Minimum	Mean	Maximum
Euclidean	Linear	51.3	68.6	82.5
Mahalanobis		45.5	59.2	77.6
Transformed Euclidean		61.4	76.9	92.3
Euclidean	Ranked	51.0	68.6	82.5
Mahalanobis		45.8	59.7	77.5
Transformed Euclidean		61.6	77.2	92.5

Table 5.3: Legendre Polynomial Hit Rate Summary.

5.2.4 The Haar Wavelet

The effect that increasing the dimensionality of the Haar wavelet feature space is shown in Figures 5.9 through Figures 5.11, which show the variation in hit rate as measured using the Euclidean, Mahalanobis, and transformed Euclidean norms.

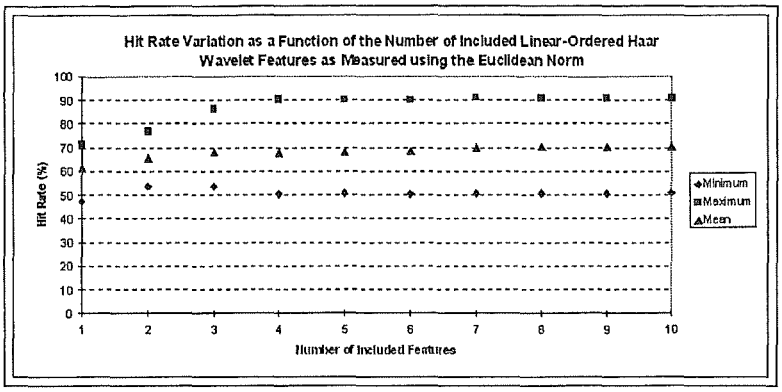

Figure 5.9: Variation in Hit Rate with the Number of Included Linear-Ordered Haar Wavelet Features as Measured using the Euclidean Norm.

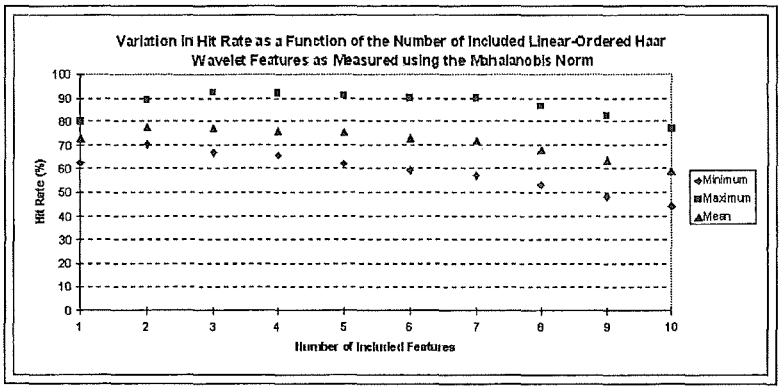

Figure 5.10: Variation in Hit Rate with the Number of Included Linear-Ordered Haar Wavelet Features as Measured using the Mahalanobis Norm.

As can be seen from Figure 5.9 through to Figure 5.11, the effect of increasing the dimensionality of the feature space is much the same as it was for the Legendre polynomial feature space. The addition of further features brings additional information to the discrimination problem that can be used to aid in the classification of a observation whose grasp class

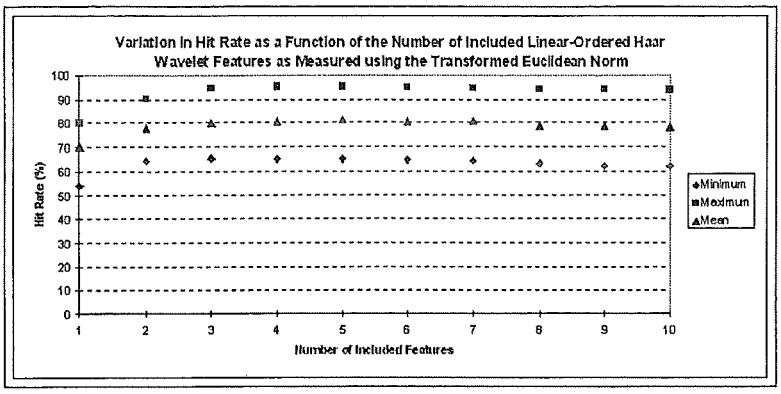

Figure 5.11: Variation in Hit Rate with the Number of Included Linear-Ordered Haar Wavelet Features as Measured using the Transformed Euclidean Norm.

identity is unknown. The effect of adding more Haar wavelet features is however slightly more pronounced than it was for the Legendre polynomial features. For instance, when using the Euclidean norm, the maximum mean hit rate achieved with the Haar wavelet features is 3% better than that achieved using the Legendre polynomials.

The effect that changing the order in which features are added to the feature space is shown in Figures 5.12 through Figures 5.14, which show the variation in hit rate as measured using the Euclidean, Mahalanobis, and transformed Euclidean norms.

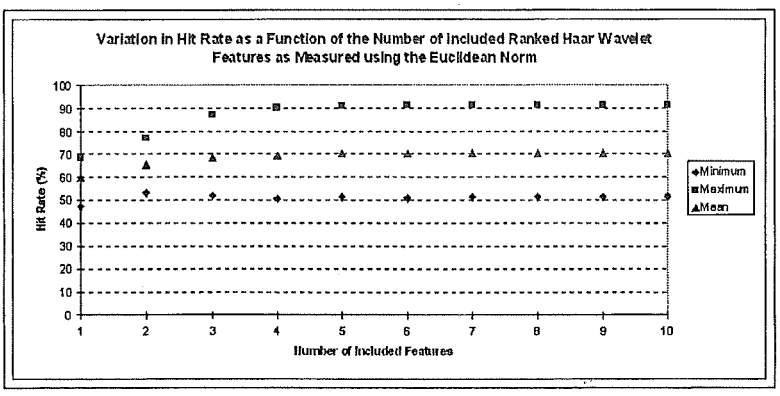

Figure 5.12: Variation in Hit Rate with the Number of Included Ranked Haar Wavelet Features as Measured using the Euclidean Norm.

The most noticeable effect of ranking the Haar wavelet features is to

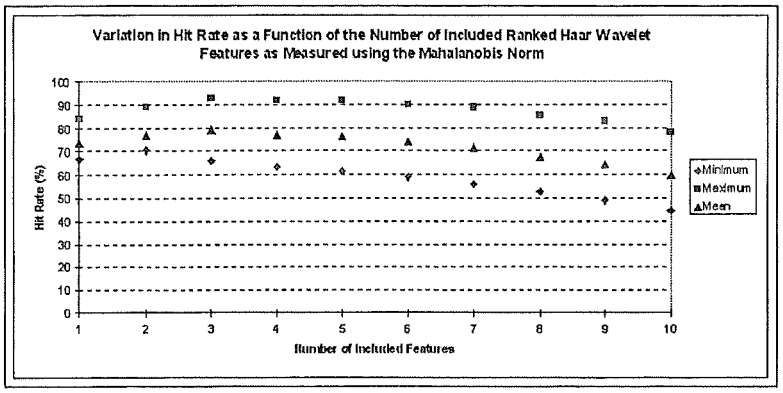

Figure 5.13: Variation in Hit Rate with the Number of Included Ranked Haar Wavelet Features as Measured using the Mahalanobis Norm.

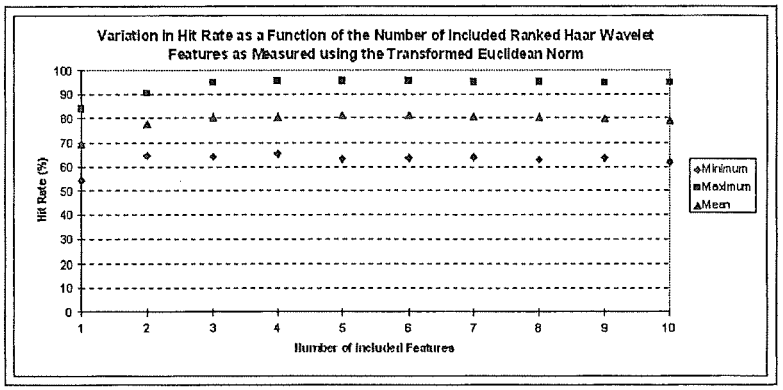

Figure 5.14: Variation in Hit Rate with the Number of Included Ranked Haar Wavelet Features as Measured using the Transformed Euclidean Norm.

reduce the number of features required to reach the maximum hit rate as measured using the three norms. This may be partially explained by examination of the ordering of the ranked Haar wavelet features. This information may be found in Table 5.4. Compared to the Legendre polynomial features, the ranking of the Haar wavelet features brings in a larger number of functions of order greater than ten. This helps to explain why it is that the ranked Haar wavelet features produce better hit rates than when they are ordered linearly. In addition to this, ranking the Haar wavelet features has the effect of slowing the rate of hit rate decline.

The hit rates obtained using the Haar wavelets are summarised in Table 5.5.

		Feature Ranking									
		1	2	3	4	5	6	7	8	9	10
Feature	16									4	1
	15						2	2	1	1	
	14									1	1
	13										1
	12						1	1	5	1	1
	11					1	1		1	1	2
	10										1
	9										
	8			2	2	5	1				
	7			2	2		3	1		1	1
	6					1		4		1	1
	5							1	2		
	4		1	3	2	1	1	1			1
	3			2	4	2	1		1		
	2	5	4	1							
	1	5	5								

Table 5.4: Haar Wavelets Feature Ranking Frequencies.

Norm	Order	Hit Rate (%)		
		Minimum	Mean	Maximum
Euclidean	Linear	50.8	70.0	90.5
Mahalanobis		44.3	58.6	76.8
Transformed Euclidean		62.2	78.5	94.0
Euclidean	Ranked	51.3	70.4	91.4
Mahalanobis		44.4	59.6	77.9
Transformed Euclidean		61.8	79.0	94.8

Table 5.5: Haar Wavelet Hit Rate Summary.

5.3 Global Dynamical Models

Two different GDM models were tested:

GDM1. This GDM model used six coefficients and a unit step as the system input.

GDM2. This GDM model used six coefficients and a tan-sigmoid function as the system input.

5.3.1 GDM1

This GDM approach used the following system model:

$$y_k = a_1 y_{k-1} + a_2 y_{k-2} + a_3 y_{k-1}^2 + a_4 yk - 1y_{k-2} + a_5 y_{k-2}^2 + a_6 \upsilon_k \qquad (5.1)$$

where y_k is the time series value at sample k, y_{k-1} and y_{k-2} are the time series values at samples $k-1$ and $k-2$ respectively, and υ_k is the value of the system input signal at time sample k. In this case, that system input signal is a unit step at $t = 0$. By implementing the GDM recursively it is possible to monitor the convergence (or otherwise) of the coefficients $a_i, i = 1, \ldots, 6$.

Given that there is no intrinsic significance in the ordering of the coefficients in Equation (5.1), the effect of both the order in which features are successively added to the feature space and the dimension of the feature space were examined at the same time. As can be seen from Figures 5.15 through to Figure 5.17, the first GDM model does not produce hit rates as high as those obtained using the orthogonal decomposition approach.

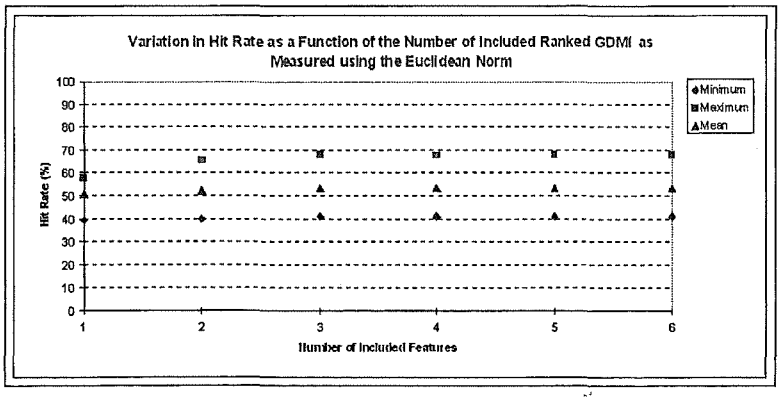

Figure 5.15: Variation in Hit Rate with the Number of Included Ranked GDM1 Features as Measured using the Euclidean Norm.

The hit rates obtained using the first GDM model are summarised in Table 5.6.

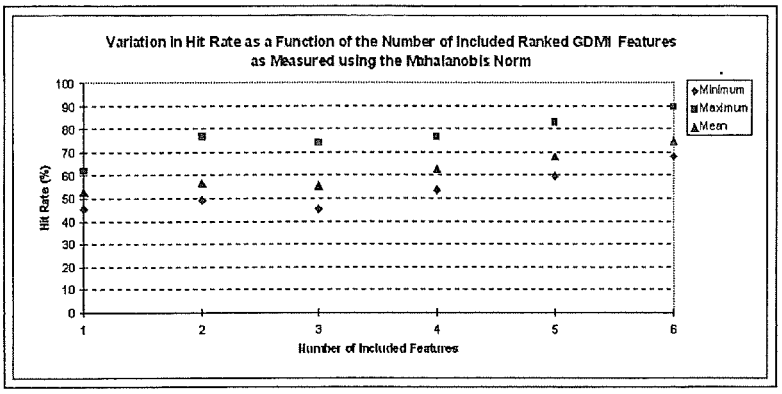

Figure 5.16: Variation in Hit Rate with the Number of Included Ranked GDM1 Features as Measured using the Mahalanobis Norm.

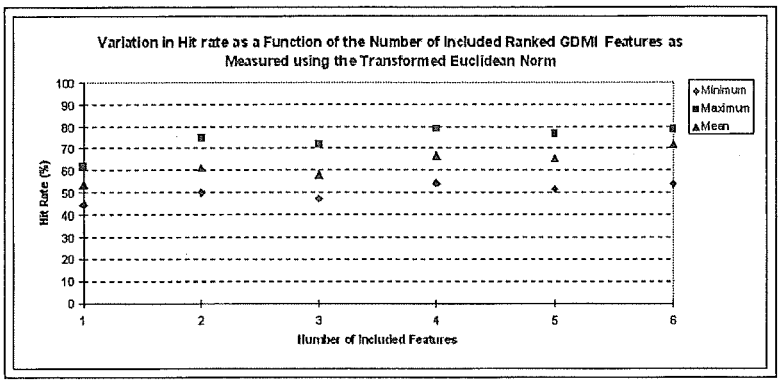

Figure 5.17: Variation in Hit Rate with the Number of Included Ranked GDM1 Features as Measured using the Transformed Euclidean Norm.

Norm	Order	Hit Rate (%)		
		Minimum	Mean	Maximum
Euclidean		41.0	53.2	68.0
Mahalanobis		68.1	74.6	89.5
Transformed Euclidean	Linear	54.6	72.3	78.5
Euclidean		41.1	53.3	68.0
Mahalanobis		68.1	74.7	89.5
Transformed Euclidean	Ranked	54.0	72.2	78.5

Table 5.6: GDM1 Hit Rate Summary.

5.3.2 GDM2

The second GDM model is the same as GDM1, except that it uses a tan sigmoid function as the system input function. This function is applied to the entire length of the EMG burst from where it first rises above the background noise by 150%. The hit rates produced using it are not greatly different to those of GDM1 and thus this GDM model must be considered inferior due to its increased computational complexity.

The hit rates obtained using the second GDM model are summarised in Table 5.7.

		Hit Rate (%)		
Norm	Order	Minimum	Mean	Maximum
Euclidean		41.2	53.2	67.8
Mahalanobis		67.5	74.9	89.0
Transformed Euclidean	Linear	53.2	70.5	78.9
Euclidean		41.2	53.2	67.5
Mahalanobis		67.5	74.0	88.9
Transformed Euclidean	Ranked	53.5	70.6	79.0

Table 5.7: GDM2 Hit Rate Summary.

5.4 EMG Histogram

Four different implementations of the EMG histogram feature extraction technique were evaluated during the course of this research. These were:

1. **HIST1.** A histogram with five 200 mV bins.
2. **HIST2.** A histogram with ten 100 mV bins.
3. **HIST3.** A histogram with 20 50 mV bins.
4. **HIST4.** A histogram with 50 20 mV bins.

The number of times that the rectified EMG signal entered each of the bins for a particular histogram implementation was recorded and these frequency counts used to parameterise the signal.

5.4.1 HIST1

The hit rates obtained using the five bin histogram are shown in Figure 5.18 through to Figure 5.20. They show the effect that increasing the dimensionality of the feature space has on the hit rate. Note that the bins of the histogram are labelled such that the first bin covers the voltage range from zero to 200 mV.

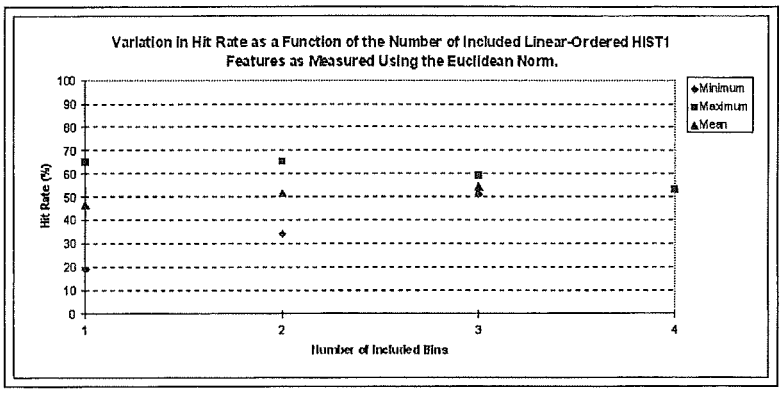

Figure 5.18: Variation in Hit Rate with the Number of Included HIST1 Bins as Measured using the Euclidean Norm.

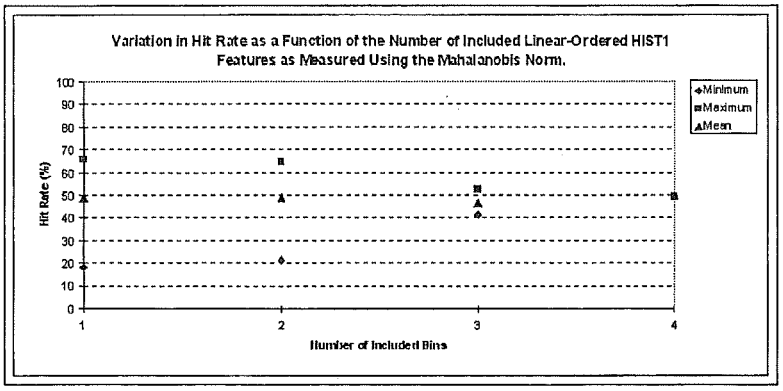

Figure 5.19: Variation in Hit Rate with the Number of Included HIST1 Bins as Measured using the Mahalanobis Norm.

As can be seen from Figure 5.18 through to Figure 5.20, the hit rates achieved using the HIST1 implementation are quite poor. This is due to the fact that there are simply not enough bins for there to be any noticeable differences between the frequency counts of signals from different grasp classes.

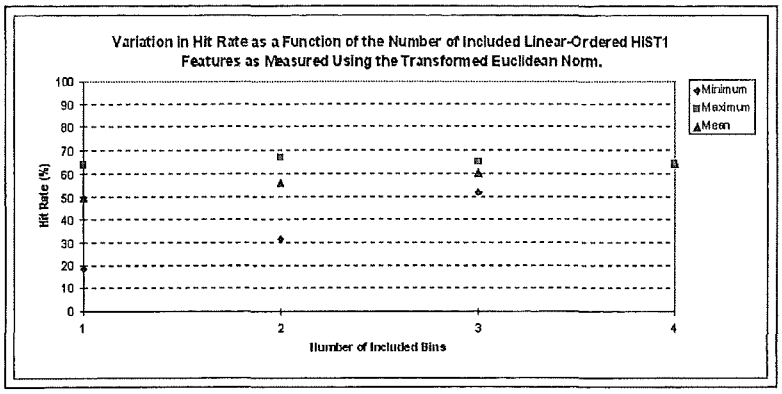

Figure 5.20: Variation in Hit Rate with the Number of Included HIST1 Bins as Measured using the Transformed Euclidean Norm.

The problem is compounded by the fact that the histogram bins have fixed voltage ranges. It was decided to use fixed voltage ranges for the bins to ensure uniformity across all data sets. As a consequence of this however, signals whose peak voltage is below 800 mV will not utilise all of the bins that are available. Examination of the independent axes of Figure 5.18 through to Figure 5.20 reveals that in fact none of the ten data sets utilised the fifth histogram bin. In addition, not all of the data sets even made full use of the first four histogram bins. Table 5.8 details just how many bins each of the ten sets of data actually used. The fact that only one data set used the first

Number of Bins	Frequency
1	2
2	5
3	2
4	1

Table 5.8: Frequency of HIST1 Bin Usage.

four bins is obvious from Figure 5.18 through to Figure 5.20 by the way that the minimum, mean, and maximum hit rates for the case of four included bins are identical. In general, it is only the first two bins that are being fully utilised, thus dramatically reducing the number of features which can be used to discriminate between the six signal classes.

Another important point to note is that there are no results available for the case where the histogram features have been ranked by their ability to provide discrimination information as was discussed in Section 5.2.2. This

is due to the fact that for each of the four EMG histogram implementations that were tested, the ranked feature order was identical to the linear feature order for all ten data sets.

5.4.2 HIST2

The second histogram technique uses ten frequency bins rather than five, and these additional degrees of freedom helped produce much better hit rates than those achieved by HIST1. This is shown in Figure 5.21 through to Figure 5.23.

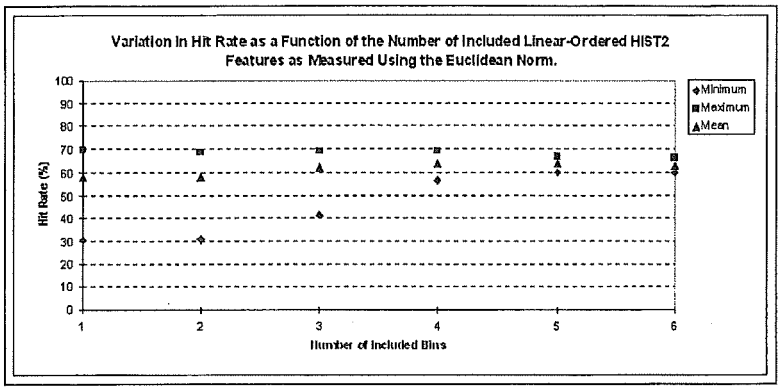

Figure 5.21: Variation in Hit Rate with the Number of Included HIST2 Bins as Measured using the Euclidean Norm.

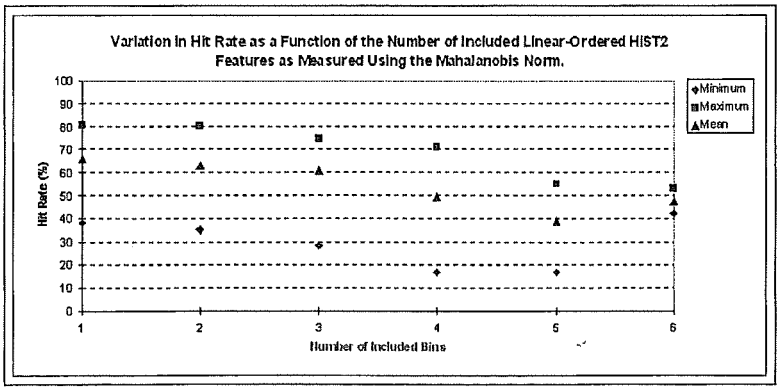

Figure 5.22: Variation in Hit Rate with the Number of Included HIST2 Bins as Measured using the Mahalanobis Norm.

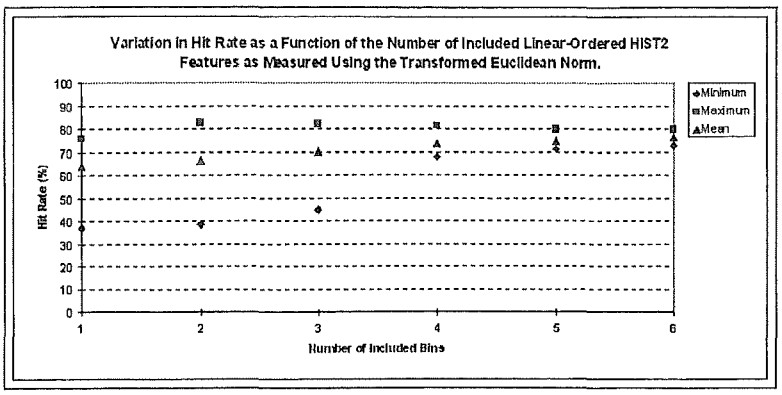

Figure 5.23: Variation in Hit Rate with the Number of Included HIST2 Bins as Measured using the Transformed Euclidean Norm.

The improvement in hit rate with the addition of an extra five bins is remarkable. Improvements of 15% are observed when using the transformed Euclidean norm. However, as with the HIST1 implementation, not all of the HIST2 bins are used. This is due to the fact that the upper histogram bin voltages are still beyond that achieved by the EMG signals. This effect must be tolerated however, as customising the bin sizes by mapping the total range of all bins to that of the signal would invalidate comparisions between grasp classes and data sets. Table 5.9 details the frequency with which particular numbers of bins are utilised by the ten sets of EMG signals. By reducing

Number of Bins	Frequency
1	0
2	2
3	3
4	2
5	1
6	2

Table 5.9: Frequency of HIST2 Bin Usage.

the voltage range spanned by each bin, the number of features by which the six signal classes may be differentiated between increases, and therefore so do the resulting hit rates. The transformed Euclidean norm is best able to take advantage of the increased degrees of freedom, and thus its hit rates show the largest improvements over the HIST1 implementation. In fact, the transformed Euclidean norms mediocre performance in the HIST1 implemen-

tation is due to the fact that there are not enough degrees of freedom for the transformed coordinate system that it produces to be much different from the original orthogonal coordinate system.

5.4.3 HIST3

Increasing the number of available bins to 20 results in only slight improvements to the mean hit rates, as shown in Figure 5.24 through to Figure 5.26.

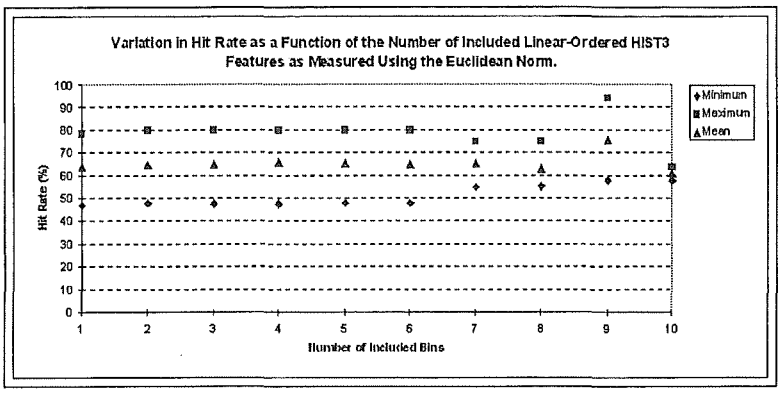

Figure 5.24: Variation in Hit Rate with the Number of Included HIST3 Bins as Measured using the Euclidean Norm.

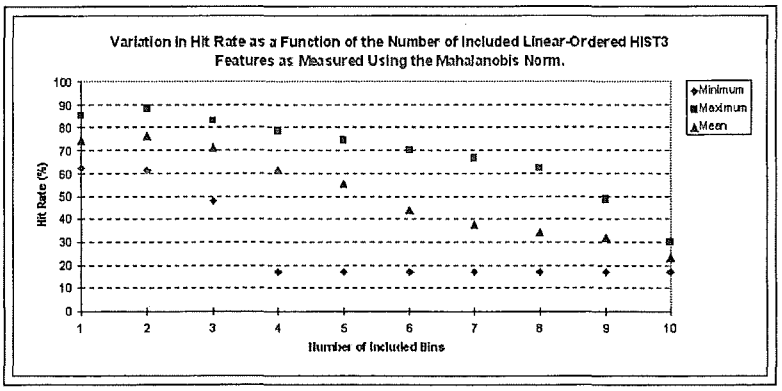

Figure 5.25: Variation in Hit Rate with the Number of Included HIST3 Bins as Measured using the Mahalanobis Norm.

These slight improvements are due to the fact that more bins are being used by the majority of the data sets, and thus it is easier to discriminate

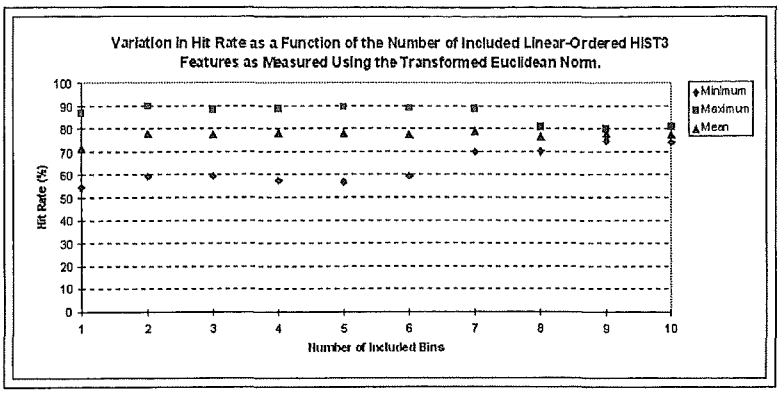

Figure 5.26: Variation in Hit Rate with the Number of Included HIST3 Bins as Measured using the Transformed Euclidean Norm.

between signal classes. This is particularly true of those data sets utilising the higher order bins. By reducing the size of the histogram bins, signals whose extreme values are different from the norm "stick out" more and are thus easier to single out in the classification procedure. Table 5.10 shows how often particular numbers of features were utilised by the ten data sets.

Number of Bins	Frequency
1	0
2	0
3	1
4	1
5	1
6	2
7	1
8	1
9	1
10	2

Table 5.10: Frequency of HIST3 Bin Usage.

5.4.4 HIST4

The fourth histogram implementation that was tested used 50 20 mV bins. This proved to be the most successful of all of the histogram techniques that

were tested, as can be seen from the hit rates in Figure 5.27 through to Figure 5.29.

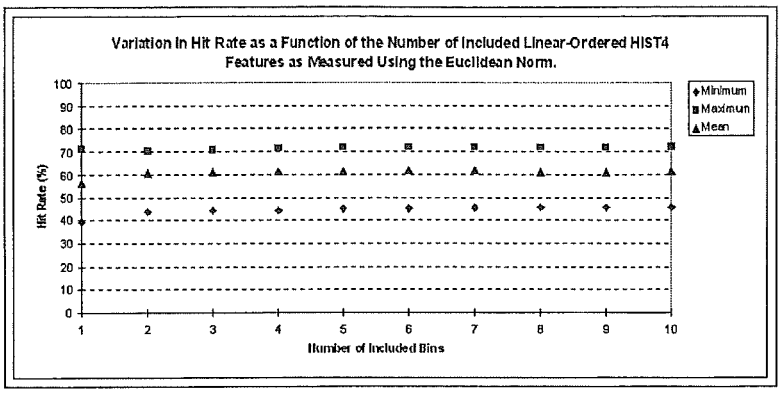

Figure 5.27: Variation in Hit Rate with the Number of Included HIST4 Bins as Measured using the Euclidean Norm.

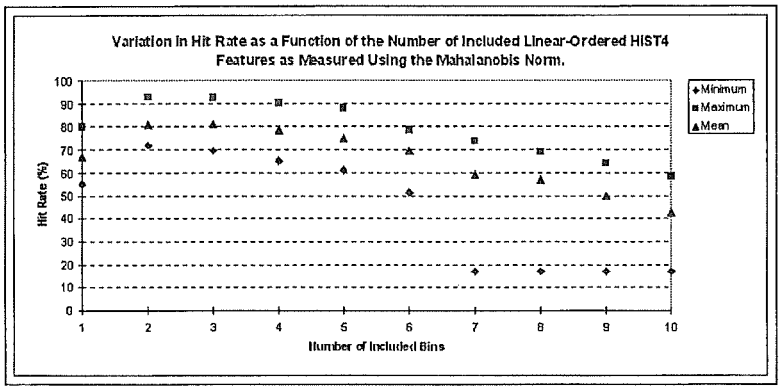

Figure 5.28: Variation in Hit Rate with the Number of Included HIST4 Bins as Measured using the Mahalanobis Norm.

Note that ten included bins are shown on the axes of Figure 5.27 through to Figure 5.29. The HIST4 implementation (along with the HIST3 implementation to a lesser extent) is the only one to actually truncate the number of bins that were used by the data sets at the limit of ten. Recall that this is done for reasons of computation time. All but one of the ten sets of clinical data utilised the first ten histogram bin frequency features in the classification scheme. With a best mean hit rate of 81%, the HIST4 implementation produced the best performance of all of the signal parameterisation techniques that have been tested. This is somewhat ironic given that the EMG

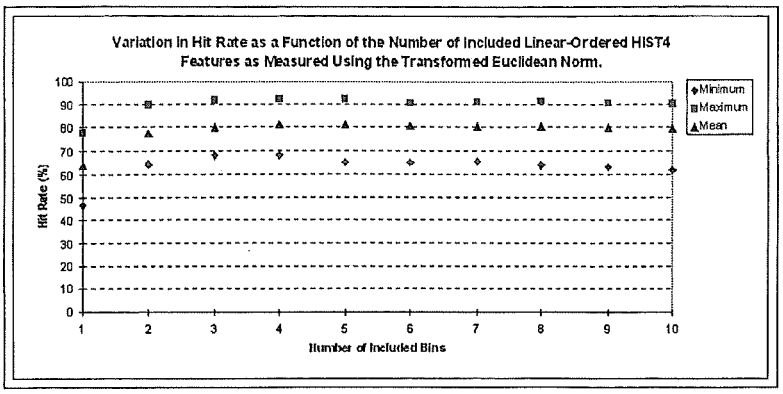

Figure 5.29: Variation in Hit Rate with the Number of Included HIST4 Bins as Measured using the Transformed Euclidean Norm.

histogram procedure is the simplest to calculate, and is the most suitable for real-time implementation. The HIST4 hit rates are also the most stable of all the EMG histogram implementations as the dimensionality of the feature space is increased. As is to be expected, it is the hit rates measured using the Mahalanobis norm that suffer the most as the number of included bins is increased. This is due to it being the most susceptible to the effect that noisy features have on the classification procedure as discussed in Section 5.2.3.

Chapter 6

Conclusions

The best mean hit rates obtained using the four signal parameterisation techniques are summarised in Table 6.1. As can be seen, of the four techniques

Parameterisation Technique	Hit Rate (%)
Envelope Maxima	75
Legendre Polynomials	77
Haar Wavelets	79
Global Dynamical Model	75
EMG Histogram	81

Table 6.1: Hit Rate Summary.

that have been evaluated during the course of this work, the EMG histogram produces the best mean hit rate across the ten data sets. Even this hit rate however is insufficient to warrant real-time implementation. With a hit rate of 81%, 19% of grasps are still being misclassified which is entirely unacceptable. Before these four techniques are dismissed out of hand however, the issue of electrode placement must be examined in more detail. As evidenced by the consistently high (in the vicinity of 95%) hit rates produced by one of the ten sets of data, there is still hope that at least one of the techniques examined here will prove to be suitable. As it is, the EMG histogram would be ideal for implementation in a real-time control scheme due to its extreme simplicity. In addition to this, the classification procedure used could easily be adapted for real-time use, given its matrix-based nature.

6.1 Future Work

There are a number of research areas that can be pursued in relation to this work. A good starting point would be a paper-based analysis of the function of each of the muscles in the forearm and hand. Determining which sets of muscles are responsible for moving the five digits is not as simple a task as it may first seem. This is because there are multiple muscles responsible for moving each digit, and that in some cases a single muscle is responsible for moving a number of digits. In addition to this, a study into the best set of grasps between which to discriminate should be undertaken. Just which grasps are selected will depend heavily on which measures of performance are used to evaluate the options. Those chosen should reflect the tasks most commonly encountered in daily life, and those to which an robotic prosthetic hand would be best suited. Armed with this knowledge, the selection of an improved set of locations from which to record EMG signals would be made much easier.

Another aspect of the problem that should be examined if there is serious intention to implement the UOC hand as a viable prosthesis is that of the method used to pick up the EMG signals. An amputee would have great trouble fitting themselves with the numerous disc electrodes and leads that were used during this research. Given that electrode position has such a bearing on the hit rates that are achieved, an alternative method of signal detection should be devised. One alternative that bears further investigation is that of permanent, subcutaneous electrodes surgically placed within the skeletal muscle from which signals are to be recorded. In addition to an improved signal-to-noise ratio due to the more intimate contact between the signal source and recording apparatus, the problem of ensuring that the recorded signals are coming from the correct muscles would no longer be an issue.

When it comes to the issue of which signal parameterisation technique to employ, the choice should not be limited to those which have been evaluated here. The discrete Fourier transform and the discrete wavelet transforms for instance hold great promise. The use of a combination of features extracted by using a number of techniques may also be worth further investigation.

The question of which classification algorithm is most suited to this task is another aspect of the problem that needs to be further researched. There are a number of alternatives to the technique used here. The k-nearest neighbour technique for example classifies an observation as belonging to the same grasp class as its k closest neighbours.

Bibliography

[Bain, 1998] Andrew R. Bain. Optimisation of a two degree of freedom finger using a genetic algorithm. Masters thesis, University of Canterbury, Christchurch, New Zealand, February 1998.

[Boca and Park, 1994] Adrian Del Boca and Dong C. Park. Myoelectric signal recognition using fuzzy clustering and artificial neural networks in real time. In *Proceedings of the 1994 IEEE Conference on Neural Networks*, pages 3098–3103, 1994.

[Dunlop and Ward, 1995] Guy R. Dunlop and Derek Kempton Ward. The kinematics of a fifteen degree of freedom six fingered hand. In *Ninth World Congress on the Theory of Machines and Mechanisms*, volume 3, pages 2249–2253, 1995.

[Farry *et al.*, 1996] Kristin A. Farry, Ian D. Walker, and Richard G. Baranuik. Myoelectric teleoperation of a complex robotic hand. *IEEE Transactions on Robotics and Automation*, 12(5):775–788, October 1996.

[Flanagan, 1998] Graham Flanagan, 1998. Personal Conversation.

[Geddes and Parker, 1975] L. A. Geddes and L. E. Parker. *Principles of Biomedical Instrumentation.* John Wiley and Sons, Inc., New York, New York, second edition, 1975.

[Griffel, 1981] D. H. Griffel. *Applied Functional Analysis.* Mathematics and its Applications. Ellis Horwood Limited, Chichester, West Sussex, 1981.

[Ito *et al.*, 1991] Koji Ito, Toshio Tsuji, Atsuo Kato, and Masami Ito. Limb-function discrimination using emg signals by neural network and application to prosthetic forearm control. In *IEEE International Joint Conference on Neural Networks*, pages 1214–1219. IEEE Press, 1991.

[Kadtke, 1995] James Kadtke. Classification of highly noisy signals using global dynamical models. *Physics Letters A*, 203:196–202, July 1995.

[Kaiser, 1994] Gerald Kaiser. *A Friendly Guide to Wavelets.* Birkhauser, Boston, Massachusetts, 1994.

[Karlik *et al.*, 1994] Bekir Karlik, Halit Pastaci, and Mehmet Korürek. Myoelectric neural networks signal analysis. In *Seventh Mediterranean Electrotechnical Conference*, pages 262–264, 1994.

[Kelly *et al.*, 1990] Michael F. Kelly, Philip A. Parker, and Robert N. Scott. The application of neural networks to myoelectric signal analysis: A preliminary study. *IEEE Transactions on Biomedical Engineering*, 37(3):221–230, March 1990.

[Kreyszig, 1988] Erwin Kreyszig. *Advanced Engineering Mathematics.* John Wiley and Sons, Inc., New York, New York, sixth edition, 1988.

[Kuribayashi *et al.*, 1992] Katutoshi Kuribayashi, Koji Okimura, and Takao Taniguchi. A discrimination system using neural network for emg-controlled prostheses. In *IEEE Workshop on Robot and Human Communication*, pages 63–68. IEEE Press, 1992.

[Ljung and Soderstrom, 1983] Lennart Ljung and Torsten Soderstrom. *Theory and Practice of Recursive Identification.* MIT Press Series in Signal Processing, Optimisation, and Control. MIT Press, Cambridge, Massachusetts, 1983.

[Magnier and Monier, 1993] Laurent Magnier and Hugues Monier. Robot hand - design of a finger. Undergraduate research project report, University of Canterbury, Christchurch, New Zealand, 1993.

[Martini, 1998] Frederic H. Martini. *Fundamentals of Anatomy and Physiology.* Prentice Hall International, Inc., Upper Saddle River, New Jersey, fourth international edition, 1998.

[Neter *et al.*, 1988] John Neter, William Wasserman, and G. A. Whitmore. *Applied Statistics.* Allyn and Bacon, Inc., Newton, Massachusetts, third edition, 1988.

[Rosheim, 1994] Mark E. Rosheim. *Robot Evolution: The Development of Anthrobotics.* John Wiley and Sons, Inc., New York, New York, first edition, 1994.

[Stringer, 1996] Scott Stringer, 1996. Personal Conversation.

[Stringer, 1998] Scott Stringer, 1998. Personal Conversation.

[Traub, 1996] Dietmar W. Traub. Weight reduction of a two degree of freedom robotic finger. Undergraduate research project report, University of Canterbury, Christchurch, New Zealand, 1996.

[Vuskovic *et al.*, 1995] Marko I. Vuskovic, A. L. Pozos, and R. Pozos. Classification of grasp modes based on electromyographic patterns of preshaping motions. In *Proceedings of the 1995 IEEE International Conference on Systems, Man, and Cybernetics*, volume 1, pages 89–95, 1995.

[Vuskovic *et al.*, 1996] Marko I. Vuskovic, J. Schmit, B. Dundon, and C. Konopka. Hierarchical discrimination of grasp modes using surface emgs. In *Proceedings of the 1996 IEEE International Conference on Robotics and Automation*, pages 2477–2483, 1996.

[Ward, 1996] Derek Kempton Ward. Design of a two degree of freedom robotic finger. Master's thesis, University of Canterbury, Christchurch, New Zealand, September 1996.

[Zardoshti-Kermani *et al.*, 1995] Mahyar Zardoshti-Kermani, Bruce C. Wheeler, Kambiz Badie, and Reza M. Hashemi. Emg feature evaluation for movement control of upper extremity prostheses. *IEEE Transactions on Rehabilitation Engineering*, 3(4):324–333, December 1995.

Appendix A

Amputation Surgical Procedure

The exact procedure followed when removing an extremity varies with the reasons for its removal. The following is a general procedure that is loosely followed when performing a mid-forearm amputation.

The procedure begins by selecting a site up the forearm at which amputation will take place. This cite is chosen so that there is enough healthy skin below it to allow for two symmetric flaps to be cut with which to cover the end of the stump. Once the skin and underlying deep fascia have been cut, as much muscle mass as possible is removed, leaving only enough to provide a covering for the bones to protect them from injury. This muscle mass has to be excised to prevent it from bunching up at the site of the amputation. It is an unfortunate fact that some of the muscle that will be removed is perfectly healthy. The muscles that remain are then transected after first having all major intramuscular vessels individually ligated. The radius and ulna will then be transected using a ring saw, although it is still not uncommon for the radius and ulna to be transected using the surgical equivalent of a pair of garden loppers. The disadvantage of this approach is hat it invariably splits the remaining bone up its length, resulting in the growth of very sharp bone spurs which then have to be surgically removed to prevent them from slicing through the overlying muscle and skin.

Once the muscle and bone are cut, the major nerves in the extremity at the point of amputation are pulled down as far as possible and cut, and all attending blood vessels are ligated at the same time. Finally, any major blood vessels that remain are individually ligated before the two flaps of skin that were shaped at the beginning of the procedure are sutured together to form a covering for the stump.

Appendix B

The Properties of L_2

As was stated in Section 4.1, the parameterisation of the time series $f_i(t), i = 1, \ldots, n$ by the maxima of their envelopes results in each grasp being characterised by an n-tuple of real numbers $(y_1, \ldots, y_n)$. The set of all such n-tuples form the real vector space $\Re^n$. A defining feature of such a set is that its members may be added together or multiplied by any real number, and the result will also be a member of the set. Stated in a more formal manner:

1. There is a rule which given any $\mathbf{x}, \mathbf{y} \epsilon \Re^n$, determines an element of $\Re^n$ called $\mathbf{x} + \mathbf{y}$ satisfying:
   (a) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ for all $\mathbf{x}, \mathbf{y} \epsilon \Re^n$
   (b) $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \epsilon \Re^n$
   (c) there is an element of $\Re^n$, called $\mathbf{0}$ such that $\mathbf{0} + \mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \epsilon \Re^n$
   (d) given any $\mathbf{x} \epsilon \Re^n$ there is an element of $\Re^n$ called $-\mathbf{x}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$
2. There is a rule which, given any $\mathbf{x} \epsilon \Re^n$ and any real number k, determines an element of $\Re^n$ called $k\mathbf{x}$, satisfying:
   (a) $k(m\mathbf{x}) = (km)\mathbf{x}$ for any numbers k, m and any $\mathbf{x} \epsilon \Re^n$
   (b) $1\mathbf{x} = \mathbf{x}$ for any $\mathbf{x} \epsilon \Re^n$
   (c) $(k + m)\mathbf{x} = k\mathbf{x} + m\mathbf{x}$ for any $\mathbf{x} \epsilon \Re^n$ and numbers k, m
   (d) $k(\mathbf{x} + \mathbf{y}) = k\mathbf{x} + k\mathbf{y}$ for any $\mathbf{x}, \mathbf{y} \epsilon \Re^n$ and any number k

These rules may also be used to define the complex vector space C by replacing the real numbers k and m by complex ones.

This concept of a vector space containing the set of all real or complex n-tuples can be expanded to include spaces containing sets of functions. There are a number of classes of function that are of great practical value when considered as a space. By abstracting various geometrical concepts into such spaces it is possible to uncover useful facts about both the space and its member functions. One example of such a space is $C[a, b]$, which is the set of all continuous real valued functions on the interval $[a, b]$, with addition and multiplication by real numbers defined in a manner analogous to those used to define addition and multiplication in the real vector space $\Re^n$ above.

The time series $f_i(t), i = 1, \ldots, n$ all belong to a modification of the space $C[a, b]$ known as $L_2[a, b]$. Recall from Section 4 that each of these n time series corresponds to a channel of recorded data for which $0 \leq t \leq T$. Therefore the time series $f_i(t), i = 1, \ldots, n$ are in fact all members of the class of functions $L_2[0, T]$. This space is the completion of $C[0, T]$ with the square integral norm. In order to understand the implications of the last statement it is necessary to examine several properties of function spaces, the first of which is the concept of a norm.

B.1 Normed Spaces

One of the most fundamental geometric concepts that needs to be abstracted from vector space is that of length. Consider the problem of using an approximate method to solve a set of equations in which the unknown is a member of a vector space. It would be greatly advantageous to be able to say when the approximation is close to the exact solution—that is, when the difference between the exact and approximate solutions is small. For a real or complex number z, the appropriate measure is the size of the modulus $|z|$; for a vector (x_1, x_2, x_3) in $\Re^3$ the appropriate measure is the length of the vector as given by Pythagoras' theorem: $(x_1^2 + x_2^2 + x_3^2)^{1/2}$. The essential idea behind the concept of length is that for every non zero vector, there is a positive number which measures its size in relation to other vectors. The abstract expression of this idea is called the norm.

A normed space is a vector space V with a given norm. A norm on a vector space V is a rule which, given any $\mathbf{x} \in V$ specifies a real number $\|\mathbf{x}\|$ (the norm of $\mathbf{x}$) such that:

1. $\|\mathbf{x}\| > 0$ if $\mathbf{x} \neq 0$ and $\|0\| = 0$
2. $\|a\mathbf{x}\| = |a| \cdot \|\mathbf{x}\|$ for all $\mathbf{x} \epsilon V$ and any scalar a
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

The same vector space can be given a variety of norms. The vector space $C[a,b]$ of real continuous functions on the interval $[a,b]$ can be turned into a normed space by defining:

$$\|f\| = \left\{ \int_a^b |f(x)|^{1/2} dx \right\}^{1/2} \tag{B.1}$$

The vector space $C[a,b]$ can be turned into a normed space different to the one in Equation B.1 by defining:

$$\|f\| = sup\{|f(x)| : a \leq x \leq b\} \tag{B.2}$$

Just which norm should be used in a particular space depends on the problem that is being solved.

The next concept that must be examined in order to understand how L_2 is related to $C[0,T]$ is that of completeness. Completeness is a term used to describe the manner in which the functions within a class converge.

B.2 Convergence

Convergence is a very important property, especially when considering problems of the approximation to the solution of various types of equation. In such problems, one has a procedure for generating a sequence of approximations which it is hoped get closer and closer to the true solution to the problem. The aim is to then to try to justify the procedure used for generating these approximations by showing the sequence converges to the true solution.

Given a sequence (x_n) of elements of a normed space N we say x_n converges to an element X of N if the norm of the difference between x_n and X can be made arbitrarily small by increasing n. This idea may be stated more formally as follows.

Let X and $x_1, x_2, \ldots$ belong to a normed space N. The series $x_1, x_2, \ldots$ converges to X if the sequence s_n converges to X, where $s_n = x_1 + \ldots + x_n$. X is then termed the sum of the series and is given by $X = \sum_1^{\inf}$.

In the space $C[a,b]$ with the sup norm (see Equation B.2) a sequence of functions converges to a function F if:

$$sup\{|f_n(x) - F(x)| : a \leq x \leq b\} \longrightarrow 0 \tag{B.3}$$

This is the condition for uniform convergence of $f_n(x)$ to $F(x)$. However, uniform convergence is not the only type of convergence that is useful.

Convergence of the normed space given in Equation B.1 means that the average (actually RMS) over the interval $[a, b]$ of the difference between $f_n(x)$ and $F(x)$ tends to zero as $n \longrightarrow \infty$. This is called convergence in the mean because it is the mean value of $f_n(x) - F(x)$ that tends to zero—the value at particular points need not always do so.

A sequence of elements of the space $C[a, b]$ may converge in the mean to a function which is not in the space. The trouble with our definition of convergence in a normed space is in a way too restrictive—it does not enable a sequence to be called convergent unless a limit element X can be produced to insert into the formal definition of convergence as stated in Section B.2. There are however many cases where a sequence may be recognised as convergent even though we cannot write down the limit ($\sum n^{-7}$ for example). It is thus useful to have a criterion for convergence which does not require a limit for the sequence in advance. For sequences of real numbers the Cauchy criterion may be used. A sequence u_n of real numbers is convergent iff for any $\epsilon > 0$ there is a number N such that $|u_n - u_m| < \epsilon$ for all $n, m\, N$. While this idea could be transferred into the language of normed spaces by replacing the $||$ by $||||$, it would not then be true in general. This is due to the fact that not all Cauchy sequences are convergent, even though all convergent sequences are Cauchy. Therefore Cauchy sequences that are convergent are given anther name.

A Cauchy sequence of elements of a normed space is a sequence x_n such that for any $\epsilon\, 0$ there is a number N such that $||x_n - x_m|| < \epsilon$ for all $m, n\, N$. A normed space is called complete if every Cauchy sequence is convergent, and incomplete otherwise. Complete spaces are more useful that incomplete spaces because a typical strategy for solving equations is to construct a sequence of approximations to a solution and then to prove it is a Cauchy sequence. In a complete space it can then be stated that the sequence converges to a member of the space under consideration.

$C[a, b]$ is incomplete with the square integral norm. It is very useful however and thus it would be greatly advantageous if it were possible to make it into a complete space. This may be done by taking any incomplete space and adding to it all the required missing elements, thus obtaining a new space ω which is complete and includes the original space as a subspace. This process is called completing the space N.

$L_2[a, b]$ denotes the completion of $C[a, b]$ with the square integral norm. Thus $L_2[a, b]$ contains all functions which are the limits of continuous functions in the sense of mean convergence. $L_2[a, b]$ may be considered as a space containing ordinary (Reimann) square integral functions together with other functions with severely discontinuous behaviour which must be included in order to make the space complete. Every function $f(x)$ such that

$\int_a^b |f(x)|^{1/2} dx$ exists belongs to $L_2[a, b]$.